



Mieux valoriser les réseaux d'épidémiosurveillance lors de l'élaboration du Bulletin de Santé du Végétal

Lucie Michel

► To cite this version:

Lucie Michel. Mieux valoriser les réseaux d'épidémiosurveillance lors de l'élaboration du Bulletin de Santé du Végétal. Agronomie. Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech), 2016. Français. NNT : 2016AGPT0026 . tel-01374036

HAL Id: tel-01374036

<https://theses.hal.science/tel-01374036>

Submitted on 29 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° 2016 AGPT 0026

Doctorat AgroParisTech

THÈSE

pour obtenir le grade de docteur délivré par

**L'Institut des Sciences et Industries
du Vivant et de l'Environnement**

(AgroParisTech)

Spécialité: Sciences agronomiques

présentée et soutenue publiquement par

Lucie MICHEL

le 30 Mai 2016

Mieux valoriser les réseaux d'épidémiosurveillance lors de l'élaboration du Bulletin de Santé du Végétal

Directeur de thèse : **David MAKOWSKI**

Co-encadrement de la thèse : **François BRUN**

Jury

Mme. Françoise LESCOURRET, Directeur de recherche, INRA Avignon
Mme. Claire NEEMA, Professeur, SupAgro Montpellier
Mme. Marie-Odile BANCAL, Maître de Conférences, INRA Grignon
M. Hervé MONOD, Directeur de recherche, INRA Jouy-en-Josas
M. François BRUN, Ingénieur, ACTA
M. David MAKOWSKI, Directeur de recherche, INRA Grignon

Rapporteur
Rapporteur
Examineur
Examineur
Co-encadrant de thèse
Directeur de thèse

N° 2016 AGPT 0026

Doctorat AgroParisTech

THÈSE

pour obtenir le grade de docteur délivré par

**L'Institut des Sciences et Industries
du Vivant et de l'Environnement**

(AgroParisTech)

Spécialité: Sciences agronomiques

présentée et soutenue publiquement par

Lucie MICHEL

le 30 Mai 2016

Mieux valoriser les réseaux d'épidémiosurveillance lors de l'élaboration du Bulletin de Santé du Végétal

Directeur de thèse : **David MAKOWSKI**

Co-encadrement de la thèse : **François BRUN**

Jury

Mme. Françoise LESCOURRET, Directeur de recherche, INRA Avignon
Mme. Claire NEEMA, Professeur, SupAgro Montpellier
Mme. Marie-Odile BANCAL, Maître de Conférences, INRA Grignon
M. Hervé MONOD, Directeur de recherche, INRA Jouy-en-Josas
M. François BRUN, Ingénieur, ACTA
M. David MAKOWSKI, Directeur de recherche, INRA Grignon

Rapporteur
Rapporteur
Examineur
Examineur
Co-encadrant de thèse
Directeur de thèse

Avant-propos



Ces travaux ont été financés par l'ACTA et l'INRA

ACTA - 149, rue de Bercy 75012 Paris

INRA - UMR Agronomie, AgroParisTech, Université Paris-Saclay,
avenue Lucien Brétignière 78850 Thiverval-Grignon, France

ainsi que par l'ANRT (convention CIFRE)



et soutenus par :



Remerciements

J'ai accepté les yeux fermés ce sujet de thèse que David a dirigé d'une main de maître et je ne pense pas m'être trompée : tu as réussi entre autre à élever ma connaissance statistique qui était proche du néant en arrivant dans l'unité, chose qui n'était pas gagnée d'avance !

Mes premiers remerciements vont ainsi à mes deux encadrants de thèse, David Makowski et François Brun. Un grand merci à vous deux de m'avoir fait confiance pour mener à bien ces travaux, ces travaux qui n'auraient pas abouti sans l'aide de votre encadrement, de nos échanges scientifiques et non scientifiques, de vos précieux conseils, de votre bonne humeur et de votre disponibilité malgré vos emploi-du-temps respectifs très chargés. Vous avez été un précieux soutien tout au long de ces trois années entre Grignon, la MNE, Toulouse ou bien même Pointe-à-Pitre.

Merci également à l'ACTA de m'avoir confié cette thèse au sein de leur projet SynOEM ainsi qu'à l'ANRT, l'un des principaux financeurs. Un immense merci à Marie-Flo pour ta bonne humeur et pour avoir géré toutes mes petites boulettes administratives ainsi qu'à Marianne et Sandrine pour tout le côté administratif et aux filles du 9^e (Violaine, Mathilde, Sonia, Émilie, Alice, Mélina) pour les quelques repas passés en votre compagnie.

Ce travail est le fruit de nombreuses collaborations. Je tiens à remercier l'ensemble des collègues de Champagne-Ardenne : Alexis Decarrier, Charlène Buridant, Vincent Bochu, Gilles Hugerot, Karim Benredjem, Mélanie Franche mais également les collègues d'Arvalis : François Piraux, Emmanuelle Gourdain, Gilles Couleaud, Danièle Simonneau et David Gouache ainsi que Jacques Veslot (ACTA); vous avez permis le bon déroulement de mon premier cas d'étude avec une finalité utile et utilisable.

Beaucoup plus au sud, en Midi-Pyrénées, je tiens à remercier Barbara Cichosz avec qui nous avons monté le partenariat pour le cas d'étude de la vigne. On continue à voyager mais un peu plus loin, outre atlantique, je remercie Catherine Abadie pour l'opportunité que tu m'as donnée à venir travailler sur vos cultures ainsi que Lucie Aurela, Marcus Héry, Christina Jacoby-Koaly pour leur confiance et pour m'avoir fait partager leurs connaissances des cultures que sont la banane et la patate douce.

Un grand merci à Philippe Vissac, Eric Gozé, André Chabert et François Piraux pour avoir pris part à mes comités de pilotage et pour les connaissances et conseils que vous avez apportés à mes travaux.

Merci à Françoise Lescourret et Claire Neema d'avoir accepté d'être rapporteurs de cette thèse et à Marie-Odile Bancal et Hervé Monod de faire partie du jury.

Je ne serais pas arrivée dans la recherche sans Marie-Hélène et Laurence de l'Unité d'Agronomie de Grignon ainsi que Catherine de l'INRA de Mirecourt, qui avec un premier stage m'ont fait découvrir ce milieu puis David a pris le relais, merci à elles. J'en oublie sûrement car en plus des quatre ans passés à l'unité, il y en a qui sont partis, arrivés, revenus, repartis, ... : merci à : Rim, Simona, Émilie, Rémy, Corentin, Michel, Alain (merci pour les publis), Joëlle, Caroline, Diana, Antoine, les ArnaudS, Anabelle (bientôt WinBUGS n'aura plus de secret pour nous), les MathieuS, Quentin, Chantal, Safia, Muriel, Elise, Raymond, Marie-Noël, Marion avec qui j'ai pu échanger plus ou moins brièvement sur mes travaux ou discuter autour d'un café et qui contribuent à la bonne entente au sein du labo. Un merci tout particulier à Mimi pour toutes tes petites attentions et les moments passés en ta compagnie, ça fait du bien ! Bon courage à Damien, Charles, Maud et Amandine pour la fin plus ou moins proche.

Tamara, un immense merci à toi pour toutes les discussions sur la vision de la vie de chercheur, de parents (un petit bisou à Jojo au passage) : tu as été d'une aide précieuse, un soutien moral et toujours de bons conseils ! En espérant avoir plusieurs choses à fêter au cours du mois de juin dans ton beau jardin !

S'il fallait que je trouve une personne pendant ces 3 ans pour accompagner mes longues journées au labo, c'était bien Clémence : parfait duo de doctorantes, portes ouvertes/portes fermées. Je pourrais en dire long mais je m'abstiendrais sur les semaines riches en émotions que l'on a pu passer ensemble surtout sur ma fin de thèse, tu as encore quelques mois à passer ici et je ne voudrai pas que les collègues prennent peur. On se sera serrées les coudes jusqu'au bout ou presque : je m'excuse de te laisser finir seule, « mais ça c'est tabou ».

Merci à Cédric, Stéphanie et Jacques pour l'intérêt qu'ils ont porté à mon métier alors que parfois je tenais le leur avec plaisir et fierté. Cette période de ma vie qui m'a permise de plancher sur mes simulations lors de longues soirées au bord de l'eau mais qui, plus sérieusement, m'a beaucoup apportée.

Merci à ceux qui ont fait que les moments hors thèse soient bénéfiques et ressourçants: les Jungles speed, team Alluets, Mèl, Alex, nicoS, matmat (à ne pas confondre avec Mat que je remercierai un peu plus tard), lolo, adèle, steph, ...

Un grand merci à mes parents et mes sœurs qui ont vécu cette aventure de plus ou moins loin au début et par la force des choses de manière très proche lors de ces 3 derniers mois. Et enfin les 3 best : Caro, Marie et Émilie (s'il te plait n'accouche pas pendant ma soutenance) qui, quoi qu'il se passe sont toujours là !

Un dernier petit clin d'œil : « Vivement avril » pour certains, pour une autre c'est tabou !

Liste des abréviations françaises

ACTA: Association de Coordination
Technique Agricole

ANRT : Association Nationale pour la
Recherche Technique

APR : Appel à Propositions de Recherche

ASSOFWI : association de producteurs de
fruits et de christophines

BSV: Bulletin de Santé du Végétal

CEFEL: Centre d'Expérimentation Fruits et
Légumes

CIFRE: Conventions Industrielles de
Formation par la REcherche

CRA: Chambre Régionale d'Agriculture

CTCS : Centre Technique de la Canne à
Sucre

DRAF: Direction Régionale de l'Agriculture
et de la Forêt

DRAAF: Direction régionale de
l'Alimentation, de l'Agriculture et de la Forêt

EA: Environnement & Agronomie

FREDON: Fédération Régionale de Défense
contre les Organismes Nuisibles

IFV : Institut Français de la Vigne et du vin

INRA: Institut National de Recherche
Agronomique

ITB: Institut Technique de la Betterave

OGM: Organisme Génétiquement Modifié

ONEMA: Organisme National de l'Eau et des
Milieux Aquatiques

PSPE : Pour et Sur le Plan Écophyto

SBT: Surveillance Biologique du Territoire

SIAFEE: Sciences et Ingénieries
Agronomiques, Forestières, de l'Eau et de
l'Environnement d'AgroParisTech

SICACFEL : Société d'Intérêts Collectifs
Agricoles Caribéenne de Fruits Et de
Légumes

SICA : Société d'Intérêts Collectifs Agricoles

SICA-LPG : SICA- Les Producteurs de
Guadeloupe

SPV : Service de la Protection des Végétaux

SRAL : Service Régional de l'Alimentation

SynOEM : Synergie, Observations, Expertise,
Modélisation

UMR: Unité Mixte de Recherche

Liste des abréviations anglaises

AIC: Akaike Information Criterion

BLUP: Best Linear Unbiased Predictors

DIC: Deviance Information Criterion

Glm: Generalised Linear Mixed Models

Glm: Generalised Linear Models

MCMC: Markov Chain Monte Carlo

RMSE: Root Mean Square Error

RMSEP: Root Mean Square Error of Prediction

SED: Stage of Evolution of the Disease

SLB: *Septoria* leaf blotch

Sommaire

Table des matières

Introduction générale	25
 Chapitre 1. Contexte historique et problématiques de la thèse.....	31
1.1 Bref historique de la protection des cultures	33
1.2 La Surveillance Biologique du Territoire (SBT).....	35
1.3 Problématiques de la thèse	40
1.3.1 Mieux valoriser les données	40
1.3.2 Échantillonnage: volonté de diminuer les coûts d'observations.....	41
1.3.3 Mise en œuvre opérationnelle par les acteurs du BSV	42
1.4 Présentation des cas d'étude	42
1.4.1 Septoriose du blé en Champagne-Ardenne	42
1.4.2 Mildiou de la vigne en Midi-Pyrénées.....	45
1.4.3 Cercosporiose jaune du bananier en Guadeloupe	48
1.4.4 Charançons de la patate douce en Guadeloupe	52
1.4.5 Besoins des acteurs	54
 Chapitre 2. Estimating the incidence of Septoria leaf blotch in wheat crops from in-season field measurements.....	57
2.1 Abstract.....	59
2.2 Introduction	59
2.3 Materials and methods.....	62
2.3.1 Data	62
2.3.2 Statistical models	64
2.3.3 Model assessment.....	67
2.4 Results.....	69
2.4.1 Estimated parameter values.....	69
2.4.2 Model selection and quality of fit	71
2.4.3 Model predictions	72
2.4.4 Posterior distribution incidence.....	75

2.5 Discussion.....	76
2.6 Acknowledgments.....	79

Chapitre 3. A framework based on generalised linear mixed models for analysing pest and disease surveys.....

3.1 Abstract.....	83
3.2 Introduction.....	83
3.3 Materials and methods.....	85
3.3.2 Model specification.....	86
3.3.3 Case studies.....	88
3.3.4 Fitting procedures and model assessment.....	90
3.4 Results.....	91
3.4.1 Septoria leaf blotch in wheat.....	91
3.4.2 Downy mildew in grapevine.....	94
3.4.3 Yellow sigatoka in banana.....	98
3.4.4 Weevils on sweet potato.....	99
3.5 Discussion.....	103
3.6 Acknowledgements.....	104

Chapitre 4. Étude de l'impact d'une réduction du nombre de parcelles observées sur la qualité des prédictions : application à la septoriose du blé en Champagne-Ardenne

4.1 Introduction.....	107
4.2 Matériel et méthodes.....	109
4.2.1 Description du dispositif actuel.....	109
4.2.2 Description des stratégies d'échantillonnage testées pour réduire le nombre de parcelles suivies.....	111
4.2.3 Mise en œuvre des stratégies « 3 strates » et « 7 strates »	111
4.2.4 Comparaison des stratégies de stratification avec de l'échantillonnage aléatoire.....	114
4.2.5 Exemple illustré de stratification et d'échantillonnage.....	114
4.3 Résultats et discussion.....	119

Chapitre 5. Co-construction et mise en œuvre opérationnelle de l'outil	127
5.1 Une thèse « appliquée » : co-construction d'un outil avec les acteurs du BSV	129
5.2 Mise en œuvre opérationnelle de l'outil	150
5.2.1 Septoriose du blé.....	150
5.2.2 Mildiou de la vigne, cercosporiose de la banane et charançons de la patate douce	152
Conclusion générale.....	153
Références bibliographiques.....	157
Annexes	171

Table des illustrations

Introduction générale

Tableau 0.1: Les 30 substances actives engendrant le retrait des autorisations de mise sur le marché de produits phytosanitaires avant le 1er février 2008.....	26
--	-----------

Chapitre 1

Figure 1.1 : Première page du BSV n°10 Grandes Cultures en Champagne-Ardenne édité le 23 avril 2015.	38
Figure 1.2 : Extrait du BSV n°10 Grandes Cultures en Champagne-Ardenne édité le 23 avril 2015 concernant la culture du blé et particulièrement la septoriose.....	39
Figure 1.3: Dispositif de l'élaboration du BSV en région Champagne-Ardenne pour la filière Grandes Cultures avec quelques détails pour la septoriose du blé.....	44
Figure 1.4: Protocole d'observation de la septoriose du blé.	45
Figure 1.5 : Cycle de développement du mildiou de la vigne.	46
Figure 1.6 : Tâche de cercosporiose jaune au stade 5 – parcelle du CIRAD Capesterre-Belle-Eau Guadeloupe.	49
Figure 1.7: Parcelle “planteur” observée dans le réseau d'épidémiosurveillance et marquage des bananiers qui sont observés lors du passage du technicien.	51
Figure 1.8: Protocole d'observation de la cercosporiose jaune du bananier.	52
Figure 1.9: Charançon de la patate douce <i>Cylas formicarius</i>	53
Figure 1.10: Schéma du piège à phéromone pour le charançon de la patate douce	54

Chapitre 2

Figure 2.1: Disease incidence data (%) collected in 2013 for the three final wheat leaves (L1,L2,L3) at one site with high risk (a) and one site with low risk (b)	64
Figure 2.2 : Disease incidence data (%) and site-specific estimates obtained with model 4F for three different sites in 2014 for leaf 3.....	69
Figure 2.3 : Residuals (observed incidence – estimated incidence) obtained with the frequentist models for leaf 2 as a function of time. a. Residuals of model 1F. b. Site-year-specific	

residuals of model 2F. c. Residuals of model 3F. d. Site-year- specific residuals of model 4F	72
Figure 2.4 : Root mean square error of prediction (RMSEP %) for four different prediction dates (May 1st, May 15th, May 30th, June 15th) and three time lags (predictions made 1 week, 2 weeks, 3 weeks in advance). a. RMSEP for leaf 1. b. RMSEP for leaf 2. c. RMSEP for leaf 3	73
Figure 2.5: Root mean square error of prediction (RMSEP %) for four different prediction dates (May 1st, May 15th, May 30th, June 15th) and three time lags (1 week, 2 weeks, 3 weeks in advance). a. RMSEP for leaf 1. b. RMSEP for leaf 2. c. RMSEP for leaf 3.....	74
Figure 2.6: Estimated posterior distribution of disease incidence for leaf 3 on May 23rd 2014, for a site-year with an observed disease incidence of 10 %.	75
Figure 2.7 : Median regional disease incidence (%) estimated for leaf 3 from 2009 to 2014.	77
 Table 2.1: Characteristics of disease incidence data for leaves 1, 2 and 3.....	63
Table 2.2: Definition of the three risk ratings used to categorise site-years	63
Table 2.3: Criteria used to evaluate the models	65
Table 2.4: Estimated parameter values / estimated posterior means of parameters and 95% confidence intervals / credibility intervals obtained for the frequentist model 4F / Bayesian model 4B.	70
 Chapitre 3	
Figure 3.1: A. Median incidence of Septoria leaf blotch estimated for two French regions in 2015 with the glmm 10septo. B. Disease incidence data and site-year-specific estimates obtained with the 10septo model for a given site-year in the Centre region in 2015..	93
Figure 3.2: Bayesian estimates versus frequentist estimates of Septoria leaf blotch incidences calculated with the glmm model 10septo, for two regions.....	94
Figure 3.3: Residuals of the six models used to analyse downy mildew incidence on grapevine as a function of time. A. Model 1vine. B. Model 2vine. C. Model 3vine. D. Model 4vine. E. Model 5vine. F. Model 6vine.....	95
Figure 3.4 : A. Median disease incidence for downy mildew estimated from 2010 to 2015 with the 6vine model. B. Disease incidence data and estimated site-specific dynamics obtained with the 6vine model for a specific vineyard in 2011, 2013 and 2014.....	96

Figure 3.5: Probability of exceeding a threshold of 250 for the yellow sigatoka SED index, by month, in Guadeloupe. A. Median estimates of probability B. Estimated values of probability for a given site-year.....	99
Figure 3.6: Estimated dynamics of weevil abundance in sweet potato crops in Guadeloupe. A. Median estimates of weevil abundance and the 95% confidence interval. B. Dynamics estimated for a specific site-year.....	101
Table 3.1: Main characteristics of the datasets.....	90
Table 3.2: Characteristics of the models tested and values of several evaluation criteria.....	97
Table 3.3: Estimated parameter values obtained with Frequentist and Bayesian methods for the selected models.....	102

Chapitre 4

Figure 4.1 : Observations d'incidence de la maladie et estimations sites-spécifiques obtenue avec le modèle pour 3 sites-années différents en 2014 pour la F3.....	110
Figure 4.2 : Exemple de stratification en 3 strates avec une sélection de 2 parcelles par strate	113
Figure 4.3 : Observations d'incidence en fonction de la date.....	115
Figure 4.4 : Incidences au 19 mai de l'ensemble des parcelles de 2014 estimées avec le modèle	116
Figure 4.5 : Observations des 15 parcelles sous-échantillonnées entre le 19 et 31 mai	117
Figure 4.6 : Incidences au 31 mai des parcelles échantillonnées et l'incidence moyenne de ces parcelles.....	118
Figure 4.7 : RMSE (%) moyens sur toutes les années des 4 stratégies d'échantillonnage pour chacune des 3 feuilles Les RMSE sont calculés pour une date de stratification au 19 mai, une date d'estimation au 18 Juin et un pourcentage de réduction des parcelles de 50%.....	120
Figure 4.8 : RMSE (%) moyens sur toutes les années pour 5 dates d'estimation pour chacune des 3 feuilles. Les RMSE sont calculés pour une date de stratification au 19 mai, la stratégie d'échantillonnage « 7 strates » et un pourcentage de réduction des parcelles de 75%.....	122
Figure 4.9 : RMSE (%) global (toutes années confondues) des 3 pourcentages de réduction de parcelles pour chacune des 3. Les RMSE sont calculés pour une date de stratification	

au 19 mai, une date d'estimation au 25 mai et la stratégie d'échantillonnage « 3 strates ».....	123
Figure 4.10 : RMSE (%) pour la F3 pour les 4 stratégies d'échantillonnage en fonction des 3 niveaux de réduction de parcelles RMSE calculées avec une date de stratification au 19 mai (SD2) et une date d'estimation au 31 mai (d5).....	124
Tableau 4.1: Nombre d'observations et de parcelles de blé observées pour chaque année et chaque feuille en Champagne-Ardenne.....	109
Tableau 4.2: RMSE (%) des 3 types de feuilles et 6 années pour chaque stratégie d'échantillonnage et % de réduction des parcelles	119
Tableau 4.3: RMSE (%) des 2 dates de stratification et pour toutes les dates d'estimation selon la date de stratification pour chaque stratégie d'échantillonnage et % de réduction des parcelles.....	121
 Chapitre 5	
Figure 5.1 : Prédictions médianes pour la feuille 3 au 23 mai 2013.....	134
Figure 5.2 : Estimation médiane obtenue en 2014 et son intervalle de confiance 50%.....	135
Figure 5.3 : Estimation selon le groupe de risque obtenue en 2014.....	136
Figure 5.4 : Rapport édité le 27 mai 2015 pour l'élaboration du BSV de la semaine 21.	146
Figure 5.5: Enquête d'utilisation du rapport Septoriose du blé -Campagne 2015- Champagne-Ardenne.....	149
Figure 5.6 : Interface web pour l'édition du rapport d'« Analyse des dynamiques épidémiologiques pour le BSV – septoriose blé tendre ».....	151
Tableau 5.1: Principaux interlocuteurs pour le cas d'étude de la septoriose du blé en Champagne-Ardenne	130
Tableau 5.2 : Échanges avec les acteurs du BSV en Champagne-Ardenne ayant permis de co-construire l'outil	132

Table des annexes

Annexe 1: Correspondence between final leaf numbers and observed leaf numbers as a function of wheat stage on the observation date	173
Annexe 2: R code used to fit a frequentist model including site-year random effects and a covariable describing the level of SLB risk (model 4F)	174
Annexe 3: WinBUGS script used to fit a Bayesian model including site-year random effects and a covariable describing the level of SLB risk (model 4B).....	175
Annexe 4: Estimated parameter values and 95 % confidence intervals obtained for the frequentist models 1F-3F, for the three leaves considered.....	176
Annexe 5: Estimated posterior means of the parameters and 95 % credibility intervals obtained for Bayesian models 1B-3B, for the three leaves considered	177
Annexe 6: Median regional disease incidence (%) estimated for leaf 2 from 2009 to 2015 in the Centre region.....	178
Annexe 7: Codes used to fit models	179
Annexe 8 : Poster présenté à l'AFPP– onzième conférence internationale sur les maladies des plantes Tours – 7 au 9 décembre 2015.....	185

Introduction générale

Le terme pesticide, dérivé du latin « caedere » (tuer) et « pestis » (fléau, calamité), est apparu dans la langue française en 1959 (INSERM 2013). Si ces produits sont très efficaces comme moyens de lutte et ont permis de forts gains de productivité et de qualité (notamment une diminution des mycotoxines), ils impactent négativement l'environnement (biodiversité terrestre et aquatique) et la santé humaine (notamment des applicateurs). Les pesticides présentent aussi une charge économique importante pour les agriculteurs dans certaines filières. La perte d'efficacité de certains pesticides liée à l'apparition de résistance est aussi une problématique émergente importante.

Aujourd'hui, les populations et les pouvoirs publics de nos pays développés prennent conscience de cette utilisation massive des produits phytosanitaires et des nombreux effets collatéraux sur l'environnement (perte de biodiversité, dégradation des milieux) et la santé humaine à toutes les échelles (formulateur, applicateur, population rurale, consommateur). Au niveau européen et national, de nombreuses substances actives ont ainsi été interdites (certaines sont présentées dans le **Tableau 1.1**) ou leurs usages ont été mieux réglementés. Au niveau national, le plan Écophyto mis en place en 2008 (MAAF 2008) et reconduit en 2015 (MAAF 2015a) a pour objectif majeur de réduire l'utilisation des pesticides en France.

Tableau 0.1: Les 30 substances actives engendrant le retrait des autorisations de mise sur le marché de produits phytosanitaires avant le 1er février 2008 (Source: Communiqué de presse, Plan Écophyto 2018, 29 janvier 2008).

Alachlore	Dinocap	Methomyl
Aldicarbe	Diuron	Molinate
Azinphos-Methyl	Endosulfan	Oxydemeton-Methyl
Azocyclotin	Fenbutatin oxyde	Paraquat
Cadusaphos	Fenpropathrine	Parathion-Methyl
Carbendazime	Fenthion	Procymidone
Carbofuran	Fenarimol	Terbufos
Chlorfenvinphos	Fluquinconazole	Tolyfluanide
Coumafène	Méthamidophos	Trifluraline
Dichlorvos	Méthidathion	Vinchlozoline

Avec ce plan, le dispositif de surveillance biologique du territoire s'est renforcé afin de mieux quantifier les problèmes sanitaires en cours de saison, étape indispensable pour pouvoir

raisonner la protection des cultures et notamment optimiser les applications de produits phytosanitaires. Le Bulletin de Santé du Végétal (BSV) s'est mis en place en 2009. Ce dispositif collectif fournit l'état de santé des cultures à l'échelle régionale sous la forme d'une édition fréquente (souvent hebdomadaire) produit à partir d'un réseau de surveillance biologique du territoire (Jullien 2013; MAAF 2015b, 2015c) comportant de nombreuses données d'observation, des prédictions issues de modèles épidémiologiques, d'expertises et des suivis biologiques en laboratoire dans certains cas. Ce bulletin permet aux agriculteurs et leurs conseillers d'avoir un diagnostic quantifié de l'état sanitaire du territoire associé à une analyse de risque chaque semaine. Ce BSV, ou sa transcription sous forme d'un bulletin technique de conseil privé, doit permettre de pouvoir mieux cibler les interventions phytosanitaires, éviter les traitements systématiques et ainsi contribuer à atteindre les objectifs du plan Écophyto. Le BSV, désormais en rythme de croisière, peut cependant être amélioré en valorisant mieux les différentes sources d'informations disponibles et en formalisant mieux leur analyse et l'intégration des différentes sources d'information.

Le volet recherche du plan Écophyto a lancé un appel à projets en 2012 « Pour et Sur le Plan Écophyto » (PSPE-1) afin de travailler sur les trois dispositifs opérationnels du plan : le réseau de surveillance biologique du territoire, le dispositif DEPHY Écophyto et le système d'indicateurs de pressions et d'impacts. Le projet SynOEM retenu à cet appel à projet (ACTA 2013), acronyme de « Synergie, Observations, Expertise, Modélisation », a pour objectif de mieux profiter de la synergie entre réseaux d'observation, expertises et modélisation pour l'élaboration du BSV, il est coordonné par l'ACTA (Association de Coordination Technique Agricole), tête de réseau des Instituts Techniques Agricoles. Ce projet est financé par l'Organisme National de l'Eau et des Milieux Aquatiques (ONEMA) dans le cadre de l'APR (Appel à Propositions de Recherche) PSPE-1 soutenu par le ministère chargé de l'agriculture.

Ma thèse s'intègre complètement dans ce projet SynOEM et propose des outils opérationnels contribuant à l'amélioration de la valorisation des nombreuses données issues du réseau d'épidémiosurveillance. Mes travaux de thèse sont menés en étroite collaboration entre l'ACTA et l'Unité Mixte de Recherche (UMR) d'Agronomie INRA (Institut National de Recherche Agronomique)/AgroParisTech dans le cadre d'une convention CIFRE (Conventions Industrielles de Formation par la Recherche). Ainsi ma thèse est également financée par une subvention de doctorat de l'Association Nationale pour la Recherche Technique (ANRT).

L'ACTA est la tête de réseau des Instituts Techniques Agricoles spécialisés par filières de productions et reconnue par les pouvoirs publics. Créés et pilotés par les agriculteurs, les Instituts Techniques Agricoles sont des organismes de recherche appliquée, d'appui technique, d'expérimentation, d'expertise, de formation et d'information. Ils ont une mission opérationnelle d'adaptation au terrain et ils jouent, à ce titre, un rôle essentiel dans la création et la diffusion du progrès technique en agriculture et dans son aval.

L'UMR d'Agronomie INRA/AgroParisTech a pour principale mission de produire des connaissances et des outils pour concevoir et évaluer des systèmes de cultures durables. Cette UMR est rattachée au département EA (Environnement & Agronomie) de l'INRA et au département SIAFEE (Sciences et Ingénieries Agronomiques, Forestières, de l'Eau et de l'Environnement) d'AgroParisTech.

Ces travaux de thèse sont présentés à travers cinq chapitres, les chapitres 2 et 3 correspondent à des articles scientifiques rédigés en anglais.

Le 1^{er} chapitre resitue la thèse dans son contexte détaillé ; la première partie de ce chapitre retrace l'histoire de la protection des cultures depuis la 1^{ère} révolution agricole jusqu'à l'élaboration en 2008 du plan national Écophyto de réduction de l'usage des produits phytosanitaires en France dans lequel s'intègre mes travaux. La seconde partie présente la problématique avec les trois questions de recherche que je me suis posée à partir du sujet de thèse, questions précisées tout au long de ces 3 ans. Ce chapitre se termine sur la présentation des quatre cas d'étude (Septoriose du blé/Mildiou de la vigne/Cercosporiose jaune de la banane/Charançons sur patate douce) qui sont la base de ces recherches.

Le 2^{ème} chapitre développe un modèle statistique permettant d'estimer l'incidence régionale et site-spécifique de la septoriose du blé en région Champagne-Ardenne à partir des observations du réseau d'épidémiosurveillance. En m'appuyant sur deux approches statistiques différentes, la statistique classique dite fréquentiste et la statistique Bayésienne, je compare plusieurs types de modèles : modèle linéaire généralisé et modèle mixte linéaire généralisé. Ce chapitre correspond à mon premier article de thèse : Michel, L., Brun, F., Piraux, F., Makowski, D., 2016. Estimating the incidence of *Septoria* leaf blotch in wheat crops from in-season field measurements. *European Journal of Plant Pathology* (published online : 19 february 2016).

Le 3^{ème} chapitre présente la méthode générale développée pour l'analyse des données issues du réseau d'épidémiosurveillance applicable à différents pathogènes et différentes cultures. Elle est basée sur l'utilisation d'un modèle mixte linéaire généralisé découlant de la méthodologie présentée dans le premier chapitre. Dans ce chapitre la méthode est appliquée sur

les quatre cas d'étude pour montrer sa grande flexibilité. Cette partie correspond à mon 2^{ème} article de thèse, actuellement en soumission chez l'éditeur de la revue *Crop Protection* : « A framework based on generalised linear mixed models for analysing pest and disease surveys ».

Après ces deux chapitres présentant des méthodes pour analyser la dynamique des maladies et ravageurs, le 4^{ème} chapitre s'intéresse à la taille du réseau d'épidémiosurveillance et à son optimisation. Différentes méthodes d'échantillonnage sont testées pour réduire le nombre de parcelles observées au sein du réseau de surveillance en cours de saison sans trop affecter la qualité de prédiction du modèle présenté dans le chapitre 1.

L'objectif du 5^{ème} et dernier chapitre de ma thèse est de présenter la mise en œuvre opérationnelle de cette démarche statistique auprès des acteurs chargés de l'élaboration du BSV dont les chambres d'agriculture et aux instituts techniques afin qu'il soit utilisé en routine lors des prochaines campagnes agricoles en région Champagne-Ardenne et région Centre.

Chapitre 1.

Contexte historique et problématiques de la thèse

1.1 Bref historique de la protection des cultures

La protection des cultures a fortement évolué au cours des derniers siècles; avant le 19^e siècle, la première révolution agricole a entraîné une modification de l'assolement des terres agricoles avec des rotations céréales-fourrages afin de nourrir les cheptels de plus en plus importants et une réaffectation de ces terres afin d'obtenir des exploitations plus compactes. Au 19^e siècle, la révolution industrielle, avec l'invention de la machine à vapeur, a entraîné une forte expansion du transport ferroviaire et maritime permettant le transport de personnes et de marchandises, cela marque un tournant dans la propagation des agents phytopathogènes et le déplacement d'espèces non-endémiques d'une région ou d'un pays à un autre (Poitout and Leclant 1986 ; Regnault-Roger 2014). Dans la deuxième partie du 19^e, nous avons vu ainsi apparaître en Europe des épidémies dévastatrices avec le mildiou de la pomme de terre en Irlande en 1845-1846 et les nombreuses maladies de la vigne telle que le mildiou, l'oïdium au début des années 1850 (Poitout and Leclant 1986 ; Jullien, J., 2013) et le Phylloxéra, fléau arrivé en France en 1863 (Pouget 1990). La protection des cultures, avant ces épidémies, était tout comme l'agriculture plutôt archaïque, dans laquelle se mêlaient les croyances religieuses, le ramassage rémunéré des divers ravageurs (campagne de hannetonnage par exemple) et quelques méthodes physiques telles que la destruction des ravageurs, l'effarouchement des oiseaux ou le brûlage des résidus de culture (Bernard 2007). Avant l'essor des produits phytosanitaires que l'on connaît aujourd'hui, le début de la lutte chimique avec des pesticides minéraux et issus de plantes pour lutter contre les maladies et ravageurs apparaît au cours de la deuxième partie du 19^e (le mildiou de la vigne sera alors traité avec de la Bouillie Bordelaise dans les années 1880). A la suite de la Seconde Guerre-Mondiale, l'agriculture mise en place au 19^e siècle va connaître une phase de modernisation (progrès technique et scientifique), pour pouvoir atteindre leur volonté d'autosuffisance alimentaire. Les recherches menées en partie par l'INRA créé en 1946 ont permis cette seconde révolution agricole. Celle-ci se traduit par l'utilisation de variétés à haut potentiel de rendement, à l'utilisation massive d'engrais chimiques et de pesticides de synthèse (fongicide, insecticides, herbicides) et au développement de la motorisation et mécanisation de l'agriculture qui devient intensive et permet d'augmenter le niveau de production agricole en France. Ces innovations agronomiques ont cependant eu des conséquences négatives importantes sur l'environnement (nitrate dans les eaux et pesticides). L'évolution que connaît actuellement l'agriculture est accompagnée par de nouvelles innovations agronomiques dans le domaine des biotechnologies végétales (OGM et sélection

assistée par la génétique), des pratiques culturelles simplifiées, et du développement de systèmes de culture bas-intrants afin de faire face au défi du XXI^e siècle : nourrir la population mondiale qui atteindra, à l'horizon 2050, 9 milliards d'individus tout en préservant l'environnement.

L'utilisation massive des pesticides est aujourd'hui remise en cause à cause de leurs effets néfastes sur la biosphère (contamination de l'air, pollution des eaux, diminution de la biodiversité) et la santé humaine (surtout pour les agriculteurs, mais aussi pour les consommateurs). Une prise de conscience de ces effets non désirables a entraîné dans la plupart des pays, la mise en place de diverses mesures pour revoir l'utilisation plus durable de ces produits phytosanitaires (Aubertot et al., 2007 ; Regnault-Roger 2014).

Des mesures réglementaires ont été mises en place dès la fin du XX^e siècle au niveau européen dans le but de préserver l'environnement (autorisation de mise sur le marché et évaluation de produits phytosanitaires) (Aubertot et al., 2007 ; Regnault-Roger 2014). En 2009, un règlement européen a vu le jour demandant d'assurer la protection de la santé humaine, animale et de l'environnement ainsi qu'une directive pour l'utilisation plus durable des pesticides. Ce règlement améliore aussi la transparence de la procédure d'évaluation des produits phytosanitaires (Regnault-Roger 2014).

La France a adopté un plan national à la suite de diverses rencontres politiques entre 2007 et 2009 dans le cadre du Grenelle Environnement avec une première version du plan Écophyto (initialement appelé Écophyto 2018) en 2008. Ce plan vise à réduire l'utilisation des pesticides de 50% dans un délai de 10 ans « si possible » et au retrait du marché des molécules les plus préoccupantes. Ce plan comporte 8 axes orientés : sur les marges de progrès des systèmes agricoles existants et la création de nouveaux systèmes de cultures minimisant l'utilisation des pesticides, autour de la communication et la formation sur la réduction et sécurisation des pesticides et sur la surveillance biologique du territoire. Ce plan est appliqué aux zones non agricoles également (parc, jardins) et inclus les spécificités des Départements d'Outre-Mer (MAAF 2008).

Les travaux de cette thèse s'inscrivent dans l'axe 5 de cette première version du plan Écophyto : « Renforcer les réseaux de surveillance des bioagresseurs et des effets indésirables de l'utilisation des pesticides ». Afin de maîtriser l'utilisation de pesticides, l'objectif de cet axe est de détecter les risques phytosanitaires émergents et de les identifier, de maintenir un bon état

phytosanitaire des cultures et enfin de détecter et identifier les effets non intentionnels des pratiques phytosanitaires pour rendre les systèmes de productions plus durables. Le rapport de Dominique Potier (député de Meurthe-et-Moselle) a établi un rapport d'évaluation pour une nouvelle version de ce plan en 2014 : « Pesticides et agro-écologie – Les champs du possible » à l'intention du premier ministre Manuel Valls (Potier 2014). Ce rapport fait le bilan des points positifs et négatifs de la première version du plan Écophyto. Une nouvelle version du plan Écophyto (Écophyto II) en 6 axes a été publiée en octobre 2015 avec la réaffirmation de la volonté de réduire l'utilisation de produits phytosanitaires de 25% à l'horizon 2020 et 50% en 2025. Dans cette nouvelle version, le dispositif de surveillance est maintenu et devra être densifié principalement dans les zones non agricoles (MAAF 2015a).

1.2 La Surveillance Biologique du Territoire (SBT)

L'arrivée du Mildiou de la vigne en France a été un élément déclencheur de l'apparition des premiers « Avertissements Agricoles » et de stations de surveillance en 1898 (Lhoste and Grison, 1989). A la suite des grandes épidémies européennes du milieu du 19^e siècle, un service de l'inspection phytosanitaire avait été mis en place (1911) en France. Il avait pour mission de contrôler l'état sanitaire des produits végétaux importés et exportés pour surveiller les organismes nuisibles émergents et absents du territoire français. En 1941, il a été transformé en Service de la Protection des Végétaux (SPV) et ses missions se sont étoffées avec le contrôle et diagnostic phytosanitaire des plantes cultivées ainsi que l'avertissement et la proposition de moyen de lutte. Ceci permettant d'anticiper et de prévenir les futures attaques pour mettre en place des stratégies phytosanitaires et éviter de traiter de façon systématique. En 1941, les Avertissements Agricoles et leurs stations passent sous la responsabilité du SPV (loi du 21 mars 1941) ; c'est le SPV puis les Directions Régionales de l'Agriculture et de la Forêt (DRAF) qui ont rédigé les Avertissements Agricoles entre 1941 et 2008 (Lhoste and Grison, 1989 ; Jullien 2013). Ces Avertissements Agricoles, envoyés aux abonnés, constituaient un outil de surveillance phytosanitaire avec une préconisation de traitement lorsque cela semblait être nécessaire d'après l'évaluation du risque. Ces Avertissements étaient basés sur des observations et des suivis biologiques réalisés en région.

En 2009, avec la première version du plan Écophyto et son axe 5, l'ensemble du dispositif de surveillance biologique du territoire et d'Avertissement Agricole a été revu en profondeur. Le

réseau épidémiologique a été étendu à l'ensemble du territoire national et des filières importantes ; les protocoles d'observation ont été harmonisés ; les observations collectées à l'échelle régionale sont centralisées et capitalisées dans une base de données ; plus d'observations sont collectées. Les Bulletins de Santé du Végétal (BSV) ont pris la suite des Avertissements Agricoles, avec un cahier des charges précisant les informations devant être publiées (pas de possibilité de préconisation de traitement) (MAAF 2015c). Ainsi, le BSV décrit chaque semaine (parfois avec une plus faible fréquence) l'état sanitaire des principaux organismes nuisibles et l'état physiologique (dont le stade phénologique) des cultures par culture dans la région considérée et présente une analyse de risque phytosanitaire en fonction du potentiel de nuisibilité des organismes. Il fait appel à des sources d'information différentes : les observations issues du réseau d'épidémiosurveillance, des données de modélisation issues de modèles épidémiologiques (lorsqu'il en existe pour le couple plante/pathogène), dans certains cas des suivis biologiques et de l'expertise. Le BSV, destiné aux agriculteurs et conseillers techniques (coopérative, Chambre d'Agriculture,...), peut être accompagné d'autres informations météorologiques et agronomiques ayant un rôle dans le développement des bioagresseurs considérés. Il peut également contenir des éléments pédagogiques comme des photographies afin de faciliter la reconnaissance des nuisibles et d'éviter les confusions avec des maladies ou insectes sans conséquence voire utiles (auxiliaires). Enfin, des éléments réglementaires demandés par les services de l'état sont également publiés dans le BSV (MAAF 2015b). Il y a un BSV par filière selon les régions et les cultures pratiquées : Grandes Cultures, viticulture, arboriculture fruitière, cultures légumières, cultures ornementales, zones non agricoles, productions spécialisées (riz, houblon, ...). Le message véhiculé par le bulletin doit inciter les agriculteurs ou leurs conseillers à se rendre dans leurs parcelles pour constater si ces parcelles sont concernées par la menace du bioagresseur mentionnée dans l'édition. L'élaboration du BSV est coordonnée par la Chambre Régionale d'Agriculture (CRA) en collaboration avec différents acteurs agricoles : Chambre d'agriculture, instituts techniques, coopératives, négoce, Fédération Régionale de Défense contre les Organismes Nuisibles (FREDON), Service Régional de l'Alimentation (SRAL), agriculteurs, celui-ci est synthétisé, validé puis publié par un animateur inter-filière de la région. Le BSV est mis à disposition gratuitement sur les sites internet de la Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt (DRAAF), Chambre d'Agriculture et des structures partenaires (MAAF 2016).

Un BSV « type » complet est composé de:

- Un titre « Bulletin de Santé du Végétal » avec sous-titre « Filière concernée » ainsi que le numéro et la date d'édition du BSV et la région considérée.
- Un paragraphe : « A retenir » qui résume pour chaque culture considérée les informations importantes
- Prévisions météorologiques de la semaine à venir
- Pour chaque culture :
 - Le dispositif d'observation (uniquement dans le premier BSV de la campagne)
 - Les stades phénologiques des parcelles du réseau
 - L'état de santé par maladies/ravageurs de la culture considérée : éléments de biologie, situation des parcelles, résultats des observations du réseau, sorties de modèle épidémiologique (lorsqu'il existe)
 - Évaluation du risque sanitaire, seuil de nuisibilité
- Mentions légales, rédaction, crédit photo, coordination et renseignements.

Les **Figures 1.1** et **1.2** présentent l'extrait d'un BSV pour la septoriose en Champagne-Ardenne.



REMARQUE : Ce bulletin est produit à partir d'observations ponctuelles réalisées du 20 au 22 avril 2015, sur parcelles agriculteur, sur 59 parcelles de colza, 52 parcelles de blé, 29 parcelles d'orge d'hiver, 31 parcelles d'orge de printemps, 14 parcelles de pois de printemps et 22 parcelles de féverole de printemps issues du réseau inter-régional IDF, Picardie et CA. S'il donne une tendance de la situation sanitaire régionale, celle-ci ne peut pas être transposée telle quelle à la parcelle.



Figure 1.1 : Première page du BSV n°10 Grandes Cultures en Champagne-Ardenne édité le 23 avril 2015.

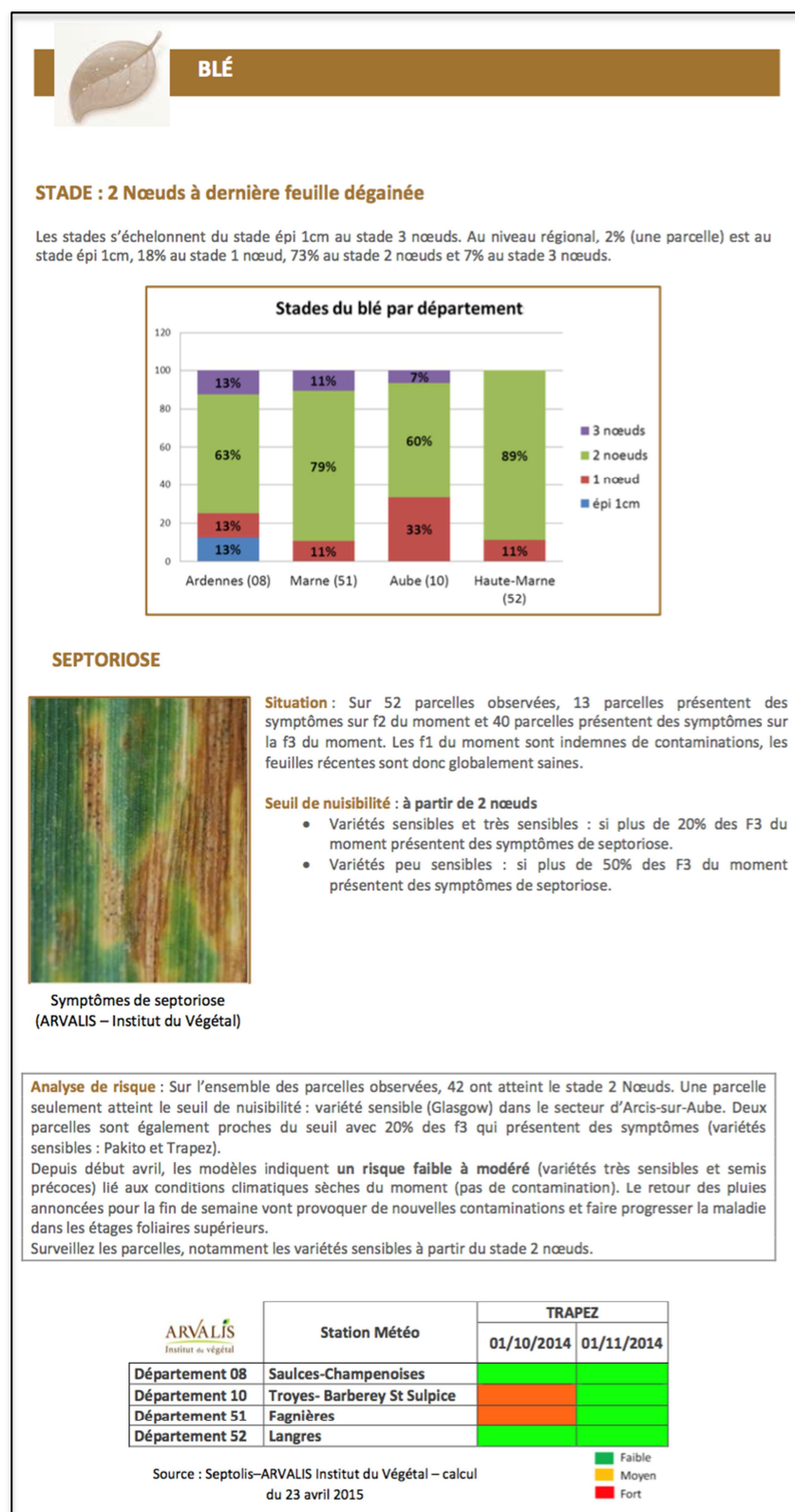


Figure 1.2 : Extrait du BSV n°10 Grandes Cultures en Champagne-Ardenne édité le 23 avril 2015 concernant la culture du blé et particulièrement la septoriose.

La première version du plan Écophyto a permis de suivre entre 2009 et 2015 plus de 15000 parcelles sur l'ensemble du territoire national avec l'aide de plus de 3500 observateurs (MAAF 2015a).

La place du BSV a été réaffirmée dans la nouvelle version du plan Écophyto (axe 1 : « Agir aujourd'hui et faire évoluer les pratiques ») avec un dispositif d'observation maintenu et densifié ainsi qu'un renfort en terme d'outil prédictif pour les maladies et agresseurs. Le BSV est actuellement fonctionnel et publié en routine mais il peut encore être amélioré, notamment en formalisant mieux le processus d'analyse et d'interprétation des données collectées via le réseau d'épidémiosurveillance. Dans ce sens, nous avons travaillé à une meilleure valorisation de ces données.

1.3 Problématiques de la thèse

Selon les régions et les filières, l'élaboration du BSV peut être différente mais elle se fait généralement dans un laps de temps réduit entre la collecte sur le terrain des observations et la publication du BSV (souvent moins de 24 heures) obligeant les acteurs à faire des choix stratégiques de rédaction. Ainsi les observations issues du réseau d'épidémiosurveillance utilisées dans le but de fournir une photographie de l'état sanitaire au niveau régional à un instant « t » sont en réalité sous-utilisées faute de temps et d'outil approprié. Dans la plupart des cas, un simple résumé est réalisé avec des statistiques descriptives : valeurs minimales, maximales, moyenne, pourcentage des parcelles atteintes.

Il n'y a pas d'élément sur la dynamique des observations permettant par exemple, l'estimation de l'évolution du pathogène pour la semaine en cours. Il n'existe pas non plus de modèle de prédiction de l'évolution du pathogène en temps réel basé sur les observations du réseau.

1.3.1 Mieux valoriser les données

Ainsi le premier objectif de cette thèse est de définir et tester des méthodes statistiques pour mieux valoriser les données issues du réseau d'épidémiosurveillance lors de l'édition du BSV. Pour répondre à ce premier objectif, les travaux reposent sur quatre cas d'étude considérant un couple culture/bioagresseur dans une région considérée.

Les quatre cas d'étude sont :

- l'élaboration du BSV pour la Septoriose du blé en région Champagne-Ardenne, qui a permis de mettre au point la méthodologie
- l'élaboration du BSV pour le mildiou de la vigne en région Midi-Pyrénées
- l'élaboration du BSV pour la cercosporiose jaune du bananier en région Guadeloupe
- l'élaboration du BSV pour le charançon de la patate douce en région Guadeloupe.

L'intérêt de se focaliser sur une région et un bioagresseur en particulier permet d'échanger avec l'ensemble des acteurs concernés en région et de mettre en place une approche globale de l'élaboration du BSV pour ce cas d'étude et ainsi proposer des améliorations en accord avec leurs besoins et de les tester.

Pour cela le travail a été décomposé en plusieurs étapes :

- la première étape pour chaque cas d'étude a été d'analyser de manière approfondie le dispositif d'élaboration du BSV afin de soulever les besoins et contraintes des acteurs pour améliorer le BSV. Cette étape a été réalisée grâce à plusieurs rencontres avec les acteurs du BSV de chaque cas d'étude qui ont exposé leur fonctionnement de collecte de données, d'analyse et de rédaction.
- la seconde étape a été la récupération des observations à travers la base de données utilisée dans le cadre du réseau d'épidémiosurveillance et la vérification de la saisie de ces observations afin d'éliminer toutes erreurs.
- l'étape d'analyse statistique a permis d'évaluer et comparer différents modèles utilisant les données récupérées pour estimer l'évolution du pathogène considéré. Lors de cette étape, plusieurs rencontres avec les acteurs ont été réalisées dans le but de présenter les avancées du projet et de co-construire le futur outil ensemble.
- la dernière étape a été de proposer un outil opérationnel aux acteurs basé sur les résultats précédents et de le tester en temps réel pendant la campagne agricole.

1.3.2 Échantillonnage: volonté de diminuer les coûts d'observations

Au cours des rencontres avec les acteurs en région Champagne-Ardenne, une problématique a été soulevée concernant la taille du réseau d'observation : une étude préalable avait été réalisée

afin de déterminer la taille du réseau pour la région Champagne-Ardenne et préconisait une moyenne de 70 parcelles à observer par an (Simonneau et al., 2011a). Cependant, nous avons réfléchi à la possibilité de réduire la taille de ce réseau en cours de saison. Notre problématique est de savoir si nous pouvons réduire le nombre de parcelles observées au cours de la saison sans trop diminuer la précision de nos prédictions. Pour répondre à cette problématique nous nous sommes basés uniquement sur le cas d'étude concernant la septoriose du blé en Champagne-Ardenne et nous avons analysé et comparé différentes méthodes d'échantillonnage.

1.3.3 Mise en œuvre opérationnelle par les acteurs du BSV

La principale finalité de la thèse est d'apporter des outils opérationnels facilitant l'élaboration du BSV par les acteurs en région. Sur la base des résultats présentés dans les premiers chapitres, nous avons proposé un outil d'estimation de l'évolution d'un pathogène pouvant être adapté à différents pathogènes. Cet outil se présente sous la forme d'un rapport produit automatiquement à partir des données du réseau d'épidémiosurveillance, pour aider à l'élaboration du BSV. L'outil est utilisable à l'aide d'une interface utilisateur disponible sur un site internet (http://appmodel.acta.asso.fr/synoem/synoem_CA/edition_rapport.php). Nous avons analysé les modalités de transfert de l'outil auprès des instituts techniques et des chambres d'agriculture.

1.4 Présentation des cas d'étude

1.4.1 Septoriose du blé en Champagne-Ardenne

1.4.1.1 Description de la maladie

La septoriose du blé *Septoria tritici* est une maladie foliaire largement répandue entraînant d'importantes pertes de rendement. Cette maladie est causée par un champignon ascomycète *Zymoseptoria tritici* (plus connu sous le nom de *Mycosphaerella graminicola*) (Eyal et al., 1983 ; King et al., 1983).

Cette maladie se développe via une reproduction sexuée et une reproduction asexuée. L'inoculum initial principal est constitué d'ascospores (forme sexuée) qui sont présents dans les pseudothèces sur les résidus de cultures. Ces ascospores sont dispersées par le vent dans l'ensemble de la culture (dispersion large) et sont responsables de l'infection primaire des

jeunes pousses de blé à l'automne. Les ascospores en se posant sur les feuilles vont germer et leur tube germinatif va pénétrer dans les stomates et se développer à l'intérieur de la feuille. Des tâches jaunes apparaissent sur les feuilles et vont se nécroser ; à la surface de la feuille, autour du stomate, le mycélium forme des pycnides (forme asexuée) qui émettent des conidies mélangées à une substance gélatineuse (cirrhe). Ces conidies sont alors disséminées par l'effet « splashing » des gouttes de pluie sur les feuilles supérieures de la même plante et sur les plantes voisines (dispersion locale). La production et la dispersion des conidies sont plus rapides que celles des ascospores, les deux reproductions coexistent pendant le cycle de la culture cependant la reproduction asexuée semble être dominante pendant la croissance de la plante (Eyal et al., 1983 ; King et al., 1983 ; Ponomarenko et al., 2011). Le pathogène survit hors de la culture principalement sous forme de pseudothèces mais aussi sous forme de pycnides sur les résidus de culture. L'infection des dernières feuilles de la plante engendrant des nécroses va entraîner une diminution de la photosynthèse et donc une perte de rendement et une diminution de la qualité des grains.

1.4.1.2 Mise en place du BSV et protocole d'observation

Le réseau BSV a été mis en place en Champagne-Ardenne en 2009. La CRA – Champagne-Ardenne est en charge du réseau d'observation et de l'élaboration du BSV. Les différents partenaires impliqués dans le dispositif de surveillance biologique du territoire sont la CRA, les chambres départementales d'agriculture, la DRAAF, les instituts techniques, la FREDON, le SRAL, les négoceurs, les coopératives et d'autres structures du secteur agricole.

L'ensemble des cultures sont représentées et regroupées en filière : la culture du blé fait parti de la filière Céréales/Protéagineux/Maïs portée par Arvalis qui élabore les protocoles d'observation, anime et rédige le BSV de cette filière. Chaque semaine, un comité de relecture du BSV est mis en place et un animateur inter-filière valide et publie le BSV (**Figure 1.3**). En 2015, 13 BSV pour la filière Céréales/Protéagineux/Maïs en Champagne-Ardenne ont été diffusés entre le 26 mars et le 18 juin 2015 ainsi qu'un BSV Bilan en fin de campagne.

Le réseau pour le blé est composé d'environ 60-90 parcelles non traitées, choisies en début de campagne agricole, observées au mieux une fois par semaine. Les parcelles observées sont réparties de manière assez homogène sur le territoire et utilisent des variétés différentes afin d'être représentatif des pratiques du territoire.

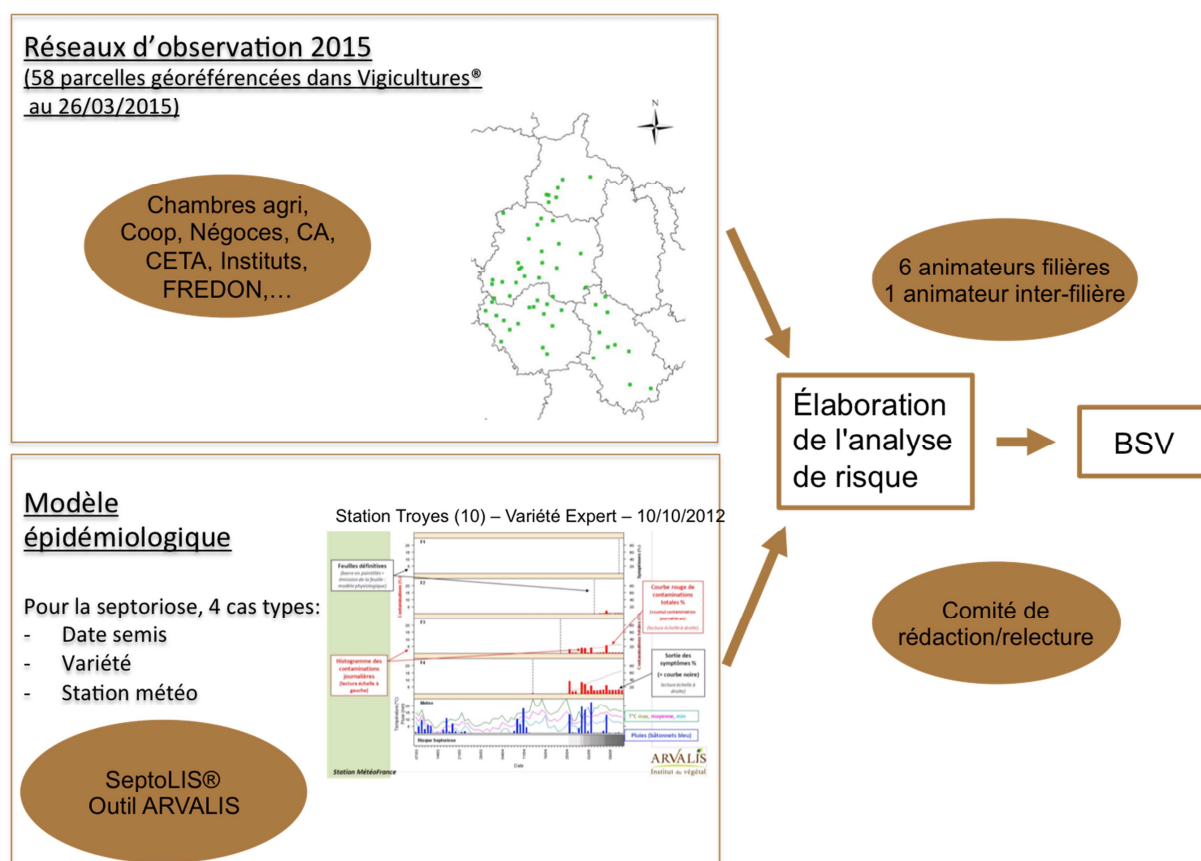


Figure 1.3: Dispositif de l'élaboration du BSV en région Champagne-Ardenne pour la filière Grandes Cultures avec quelques détails pour la septoriose du blé (source : carte du réseau issue du BSV Champagne-Ardenne n°6 – 26 mars 2015 - Grandes Cultures et sortie du modèle Septolis issue du BSV Champagne-Ardenne n°19 – 07 mai 2013 – Grandes Cultures).

En début de campagne, les caractéristiques de chaque parcelle sont renseignées : nom de la parcelle, géolocalisation (souvent précision au niveau de la commune et pas de la parcelle), date de semis, variété. Puis chaque semaine, les observations sont réalisées de façon assez souple (du lundi au mercredi principalement) mais doivent être saisies dans la base de données Vigicultures® (base de données nationale pour les Grandes Cultures proposée par Arvalis, Terres Inovia et ITB (Institut Technique de la Betterave)), avant le mercredi 14h car la rédaction du BSV est réalisée entre le mercredi après-midi et le jeudi matin. Le comité de relecture se réunit au téléphone le jeudi en début d'après-midi afin d'obtenir un BSV final. Puis, il l'envoie à l'animateur inter-filière qui agglomère tous les messages, le met en forme et le met en ligne dès le jeudi soir sur le site de la CRA et la DRAAF.

Les observations pour la septoriose du blé se font du mois de mars à juin sur des témoins non traités des parcelles du réseau (taille de 18m sur 22m). Elles répondent au protocole harmonisé

de Vigicultures®, 20 plantes sont prélevées aléatoirement dans le témoin et les 3 dernières feuilles déployées de chaque plante sont inspectées visuellement de la présence ou l'absence de la septoriose. Une note d'incidence sur 20 est alors attribuée pour chaque feuille (F1, F2 et F3) chaque parcelle (0 lorsque aucune feuille est atteinte et 20 lorsque l'ensemble des feuilles sont atteintes) (Simonneau et al., 2011b). Ce protocole d'observation est illustré par la **Figure 1.4**.

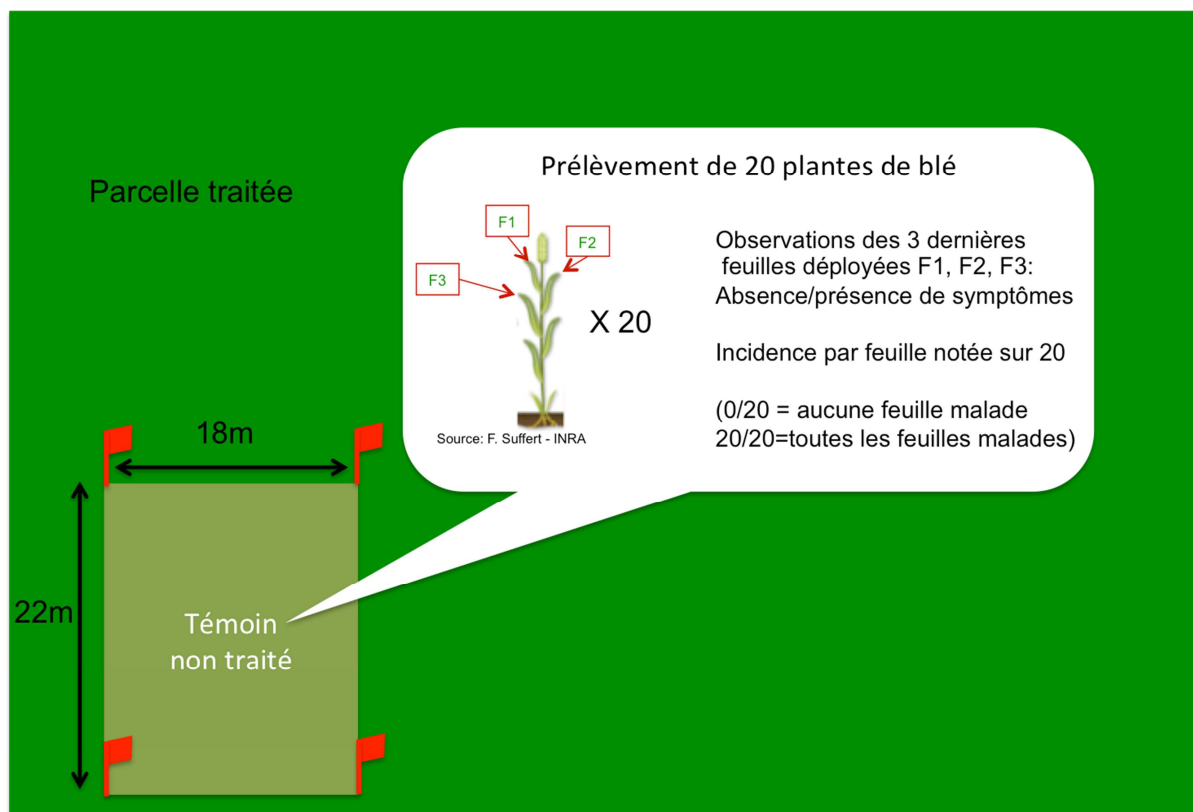


Figure 1.4: Protocole d'observation de la septoriose du blé.

1.4.2 Mildiou de la vigne en Midi-Pyrénées

1.4.2.1 Description de la maladie

Le mildiou de la vigne est une maladie causée par un oomycète *Plasmopara viticola*. Les organes touchés par le mildiou sont tous les tissus verts de la vigne : les feuilles avec l'apparition de taches jaunâtres en début de contamination et le dessèchement de celles-ci mais aussi les grappes, les sarments et les rameaux.

La **Figure 1.5** illustre le cycle de cet oomycète qui se conserve dans les feuilles mortes tombées au sol sous la forme d'oospores. Selon des conditions particulières climatiques (pluviométrie et température), ces oospores vont mûrir pendant l'hiver et germer au printemps en formant des sporanges. Ces sporanges vont libérer des zoospores qui vont contaminer les organes de la vigne via les éclaboussures de pluie. Les zoospores vont s'introduire dans le limbe de la feuille par les stomates, un mycélium va se former s'alimentant des réserves de la feuille sur laquelle apparaît une tache blanche sur sa surface inférieure (« tâche d'huile ») : c'est la contamination primaire. A la suite de cette infection primaire, des sporanges vont sortir des stomates (feutrage blanc sur la surface inférieure des feuilles) ce qui va entraîner une libération de zoospores ainsi que leur dispersion et germination pour aboutir à l'infection d'un nouveau tissu de vigne : c'est l'infection secondaire (reproduction asexuée). De même, à la suite de l'infection primaire, la reproduction sexuée permet la formation d'oospores dans le tissu à partir du mycélium qui seront libérés pour une future maturation et germination (conservation hivernale) (Gessler et al., 2011). Cette maladie génère de fortes chutes de rendement (perte de récolte), des problèmes de qualité des vins (modification de l'arôme et du goût) et un affaiblissement des ceps (Jermini et al., 2010).

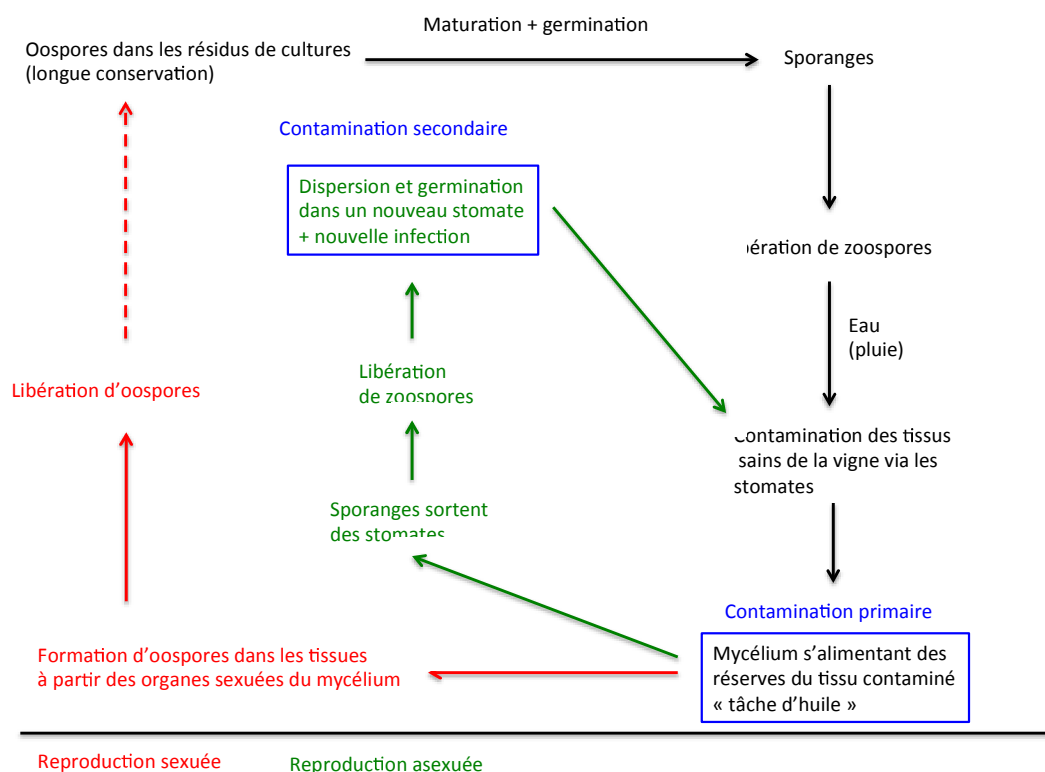


Figure 1.5 : Cycle de développement du mildiou de la vigne.

1.4.2.2 Mise en place du BSV et protocole d'observation

Comme dans les autres régions, c'est la CRA (Midi-Pyrénées) qui a la responsabilité de l'élaboration des BSV et la coordination du dispositif de SBT pour l'ensemble des filières. Pour la viticulture, les acteurs du BSV sont : la CRA-Midi-Pyrénées (rôle d'animateur inter-filière), les chambres d'agriculture départementales, les coopératives, les négoce, les organisations de producteurs qui gèrent le réseau d'observation, la SRAL, le Centre d'Expérimentation Fruits et Légumes (CEFEL), l'Institut Français de la Vigne et du vin (IFV) pôle Sud-Ouest, qui s'occupent des modèles épidémiologiques et la FREDON et le CEFEL qui ont en charge les suivis biologiques. Tout comme le BSV Grandes Cultures en Champagne-Ardenne, l'élaboration de l'analyse de risque épidémiologique est effectuée par les animateurs filières et un comité de rédaction afin de rédiger le BSV.

Particularité régionale, la région Midi-Pyrénées a choisi de publier un BSV pour chaque vignoble dans la mesure où ils ont des cépages et des caractéristiques pédo-climatiques bien distincts. Il y a donc 8 éditions BSV de rédigées chaque semaine : Fronton, Gaillac, Cahors-Lot, Gascogne-Saint-Mont-Madiran, Tarn-et-Garonne, Aveyron, Limousin (accord avec la région voisine), Raisin de table (pour plus d'échange entre les deux filières).

En 2015, un total de 134 bulletins a été publié durant la campagne avec un maximum de 20 BSV pour chacune des éditions ainsi qu'un bulletin bilan fin de campagne pour chacune d'elle.

Dans le cas du mildiou de la vigne, la publication du BSV s'effectue le mardi de chaque semaine, ainsi, les différentes informations issues du réseau d'observation, des modèles épidémiologiques et des suivis biologiques sont collectées et vérifiées la veille (lundi). Le comité de rédaction se réunit au téléphone le mardi matin.

Le réseau d'épidémiosurveillance est composé de parcelles agriculteurs non traitées (témoin non traité - TNT) et traitées (référence agriculteur – REF), celles-ci sont géolocalisées et différentes caractéristiques sont référencées pour chacune d'elles (comme le cépage, âge de la plantation,...). Le réseau est défini au début de chaque campagne, puis chacune des parcelles est suivie tout au long de la campagne au mieux une fois par semaine si possible à jour fixe le lundi. Le témoin observé doit comporter au minimum 50 ceps, l'ensemble des ceps est alors observé ; l'observateur renseigne la date, le stade, le nombre de ceps observés, la fréquence de ceps

porteurs des symptômes de mildiou, mais aussi la fréquence et l'intensité sur feuilles et sur grappes des symptômes. D'autres maladies et ravageurs sont observés sur ces mêmes parcelles. Ces observations sont saisies dans la base de données nationale EPICURE (IFV 2010).

1.4.3 Cercosporiose jaune du bananier en Guadeloupe

1.4.3.1 Description de la maladie

La cercosporiose jaune est une maladie foliaire du bananier causée par un champignon ascomycète *Mycosphaerella musicola* R. Leach ex J.L. Mulder, elle est appelée aussi « maladie de Sigatoka » (à ne pas confondre avec la cercosporiose noire ou maladie des raies noires, « black sigatoka », causé par *Mycosphaerella fijiensis*). Identifiée pour la première fois à Java en 1902, elle est présente dans la plupart des zones de production de banane depuis 1962 (Meredith 1970). Il existe cinq stades d'évolution pour cette maladie : le premier stade (stade 1) sont des minuscules points jaunes/vert-clair inférieurs à 1mm sur la face supérieure des feuilles (difficilement observables), le deuxième stade (stade 2) les tirets s'allongent pour atteindre une taille de 1 à 5mm, le troisième stade (stade 3) correspond à l'élargissement des tirets et leur centre devient de couleur brun rouille, au quatrième stade (stade 4), les lésions brunes entourées d'un halo jaune mesurent entre 12 et 15mm de long et 2 et 5mm de large, le cinquième et dernier stade d'évolution (stade 5) correspond au dessèchement du centre des lésions nécrotiques qui devient gris entouré d'un anneau noir et d'un halo jaune (**Figure 1.6**) (Lassoudière 2007).



Figure 1.6 : Tâche de cercosporiose jaune (longues tâches jaunes nécrosées) au stade 5 – parcelle du CIRAD Capesterre-Belle-Eau Guadeloupe.

La réduction de sa surface foliaire due aux lésions nécrotiques entraîne une diminution de la photosynthèse, cette maladie est ainsi responsable d'une perte de rendement de la récolte brute et une accélération de la maturation des fruits à l'origine d'une diminution du rendement de la récolte exportable (Fouré 1994).

Notre troisième cas d'étude traite de cette maladie sur les bananes *Musa spp.* de variété Cavendish (banane dessert) en Guadeloupe qui sont principalement exportées.

La cercosporiose se propage par l'intermédiaire de deux vecteurs : l'eau et le vent. D'une part, sous l'action de précipitations, les conidies mûres sont expulsées par les conidiophores se développant sur les lésions à partir du stade 4 (reproduction asexuée). Ces spores sont alors transportées (principalement verticalement) par les eaux de pluie des feuilles supérieures vers les feuilles inférieures de la plante (contamination de la même plante) et les rejets. D'autre part, sous l'action de la rosée ou de précipitations, les périthèces développés à l'intérieur des lésions nécrosées du stade 5 libèrent des ascospores (reproduction sexuée). Ces ascospores sont transportées dans toutes les directions par le vent et diffuse la maladie sur une plus grande distance (contamination de la même plante et des plantes voisines) (Fouré 1994 ; Lassoudière 2007).

1.4.3.2 Mise en place du BSV et protocole d'observation

En Guadeloupe, le plan Écophyto a été mis en œuvre progressivement à partir de 2010. Le réseau d'épidémiosurveillance est géré par six organismes : la Chambre d'Agriculture, la SICA-LPG (Société d'Intérêts Collectifs Agricoles - Les Producteurs de Guadeloupe), le CTCS de Guadeloupe (Centre Technique de la Canne à Sucre), l'ASSOFWI (association de producteurs de fruits et de christophines), la SICACFEL (SICA CARibéenne de Fruits Et de Légumes) et la FREDON Guadeloupe.

Les premiers BSV « Canne à Sucre et Banane » sont apparus en 2011. C'est seulement en 2013-2014 que chaque filière dispose de son propre BSV : arboriculture (agrumes), vivrière (patate douce, igname), horticulture (anthurium), canne-a-sucre, banane et maraichage. En 2015, 42 BSV ont été rédigés : 7 en arboriculture, 4 en cultures vivrières, 4 en horticulture, 10 en culture de canne à sucre, 11 en culture de banane et 6 en maraichage. Ce qui porte à 140 BSV depuis le début de leurs éditions en Guadeloupe. A la différence des cultures en métropole, il n'existe aucun protocole national pour les cultures tropicales, les protocoles sont cependant harmonisés et établis à l'aide d'instituts techniques et scientifiques. Ainsi un BSV « Banane » apparaît environ tous les mois.

Le réseau pour la filière banane est actif depuis 2004 et piloté par l'EURL SERVIPROBAN et la SICA-LPG. Pour la cercosporiose jaune du bananier en Guadeloupe : les observations se déroulent sur toute l'année avec une fréquence hebdomadaire. Les observations se font sur un échantillon de 10 bananiers non fleuris adultes (que l'on garde tout au long de la campagne d'observation) dans chaque parcelle observée (**Figure 1.7**) et elles s'effectuent sur le cigare et les 5 plus jeunes feuilles : pour chaque feuille une note d'intensité de symptômes est attribuée de 0 à 5 (correspondant aux 5 stades d'évolution de la maladie). Les parcelles sont des parcelles agriculteurs traitées.



Figure 1.7: Parcelle “planteur” observée dans le réseau d’épidémiosurveillance et marquage des bananiers qui sont observés lors du passage du technicien.

Une note d’État d’Évolution (EE) détermine la contamination de la parcelle ; chaque semaine, ce EE prend en compte le nombre total de feuilles, le stade de développement du cigare et le stade de développement de la maladie sur les cinq plus jeunes feuilles ainsi que le cigare du bananier (Lassoudière 2007 ; Ganry 2008). Le protocole d’observation est illustré par la **Figure 1.8**. Si

l'EE est compris entre 0 et 250 exclu, la contamination est considérée comme faible, entre 250 et 500 inclus comme moyenne et si elle est supérieure à 500 alors elle est considérée comme forte.

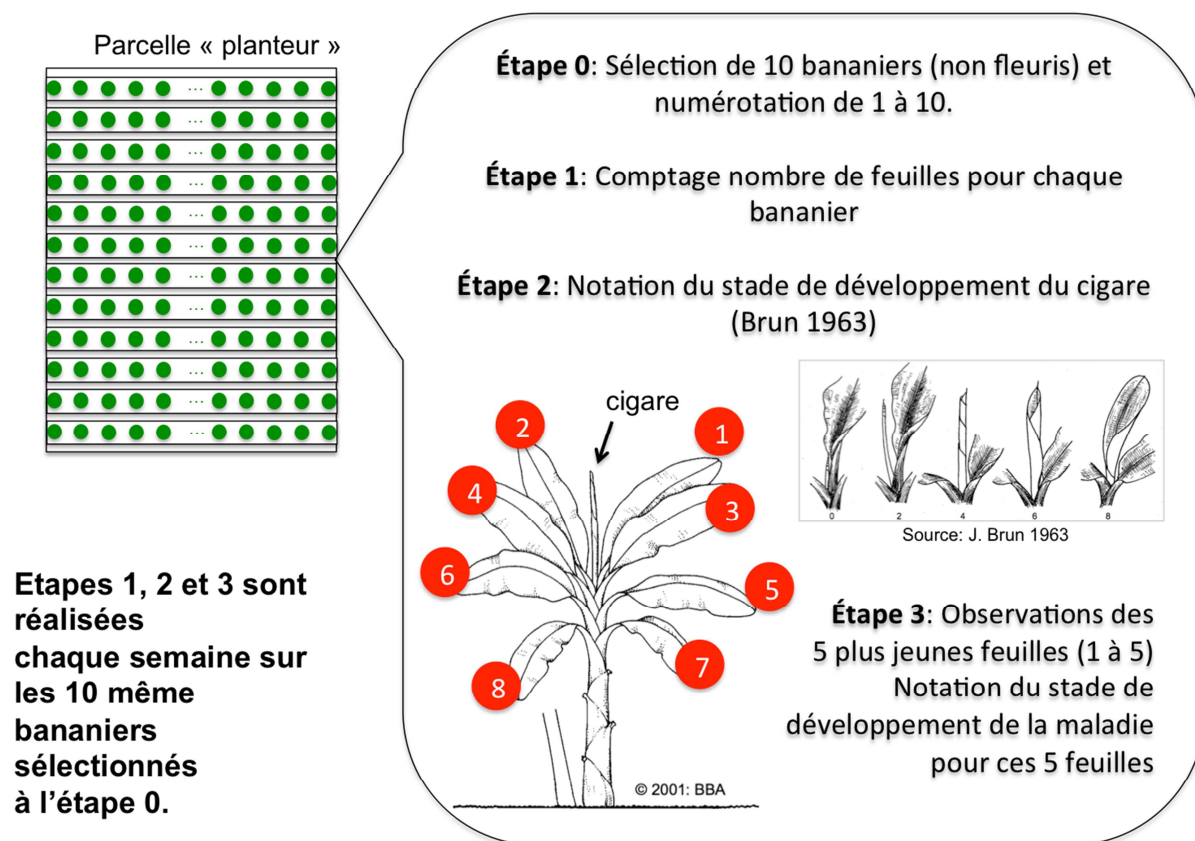


Figure 1.8: Protocole d'observation de la cercosporiose jaune du bananier.

1.4.4 Charançons de la patate douce en Guadeloupe

1.4.4.1 Description de la maladie

Le charançon est un ravageur des cultures, connu sous le nom de *Cylas formicarius* Fabricius et appartenant à la famille des Brentidae (Ordre: Coleoptera). Cet insecte (**Figure 1.9**) a été découvert en Guadeloupe en 1999 et est un important ravageur de la culture de patate douce (*Ipomoea batatas* Linnaeus), il est également présent sur une mauvaise herbe du littoral : la patate bord de mer (*Ipomoea pes-caprae*). En Guadeloupe, les parcelles attaquées sont principalement localisées sur la Grande-Terre et le nord de la Basse-Terre. Les femelles charançons perforent les tiges et tubercules de patate douces afin d'y pondre ses œufs. Une fois

écloses, les larves vont alors creuser les tubercules ce qui va les déprécier et les rendre impropre à la consommation ; les dégâts causés peuvent entraîner parfois la destruction complète de la récolte (Denon and Mauléon, 2004 ; Mauléon and Denon 2008).



Figure 1.9: Charançon de la patate douce *Cylas formicarius* (source : BSV région Guadeloupe n°2 du 23 mars 2015 – Culture vivrière).

1.4.4.2 Mise en place du BSV et protocole d'observation

Tout comme pour la filière banane, les premiers BSV sont apparus en 2011 pour les cultures vivrières (patate douce et igname).

Le réseau de suivi épidémiologique pour le charançon de la patate douce a été mis en place à la suite du plan ÉcoPhyto et est piloté par la SICACFEL. Les observations sont réalisées grâce aux piégeages des adultes mâles avec un piège à phéromone. Les parcelles sont des parcelles agriculteurs non traités (il n'existe pas de traitement pour le charançon). Six zones géographiques ayant été définies pour le suivi, il y a au minimum une parcelle observée dans chaque zone. Un piège (**Figure 1.10**) par parcelle est mis en place un mois après le début de la plantation jusqu'à la fin de cycle de la culture, le comptage du nombre de charançons piégés se fait toutes les 1 à 3 semaines au mieux (Champoiseau et al., 2011). Ainsi un BSV est publié dans le meilleur des cas tous les mois, selon la fréquence des suivis.

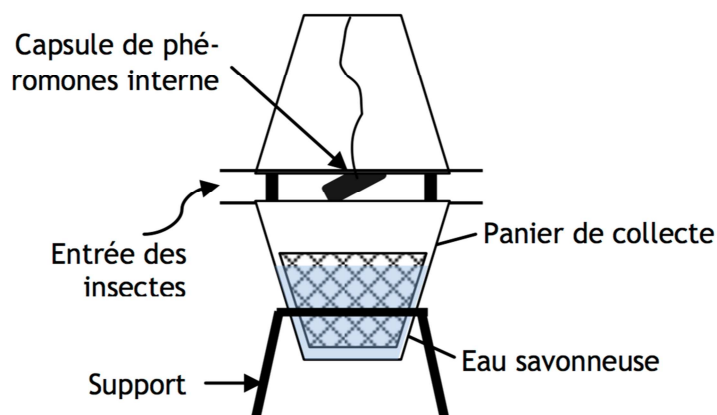


Figure 1.10: Schéma du piège à phéromone pour le charançon de la patate douce (Source : BSV région Guadeloupe n°2 juillet 2011 - Diversification végétale).

1.4.5 Besoins des acteurs

D'après plusieurs rencontres au début de ma thèse pour l'ensemble des cas d'étude, le constat réalisé par rapport à l'élaboration des BSV en région est (i) la nécessité de pouvoir faire un traitement rapide des informations recueillies du fait du peu de temps disponible entre la saisie des observations dans les bases de données d'épidémiosurveillance et la rédaction du BSV, et (ii) le besoin d'améliorer la valorisation des observations collectées en cours de campagne pour la rédaction du BSV ainsi qu'en fin de campagne. Les données sont actuellement sous-valorisées dans la mesure où de simples indicateurs descriptifs sont utilisés, comme la moyenne ou le pourcentage de parcelles ayant atteint un niveau d'incidence donnée. Il n'y a pas d'analyse formelle de ces données. Les données publiées dans le BSV sont celles de la semaine et il n'y a pas d'analyse de la dynamique de la maladie/ravageur, ni de repères par rapport aux années passées.

En réponses à ces besoins : nous proposons de décrire l'évolution des observations semaine par semaine, du début à la fin de la campagne, et de prédire la dynamique des bioagresseurs. Notre but est d'apporter aux acteurs une information pertinente sur la dynamique des bioagresseurs afin qu'ils adaptent leurs stratégies de traitement.

Dans les deux chapitres suivants, je présente les travaux réalisés afin de mettre en œuvre ces propositions pour la septoriose du blé en région Champagne-Ardenne (Chapitre 2) et je mets en avant le caractère générique de la méthode en m'appuyant sur les trois autres cas d'étude

(mildiou de la vigne, cercosporiose jaune du bananier, charançons de la patate douce) (Chapitre 3). En parallèle de ces travaux, à la demande des acteurs de Champagne-Ardenne, nous avons évalué l'impact d'une réduction de la taille de l'échantillon de parcelles observées sur la qualité des prédictions (chapitre 4). Le chapitre 5 met en avant les démarches de transfert de la méthode aux instituts techniques pour qu'elle soit utilisée en temps réel en région.

Chapitre 2.

Estimating the incidence of Septoria leaf blotch in wheat crops from in-season field measurements

Ce chapitre correspond à un article publié (Michel, L., Brun, F., Piraux, F., Makowski, D., 2016. Estimating the incidence of Septoria leaf blotch in wheat crops from in-season field measurements. European Journal of Plant Pathology (published online 19 february 2016)).

Lucie Michel¹, François Brun², François Piraux³, David Makowski⁴

(1) ACTA, INRA, UMR Agronomie, AgroParisTech, Université Paris-Saclay, 78850 Thiverval Grignon, France

(2) ACTA, INRA, UMR AGIR, F-31326 Castanet Tolosan, France

(3) ARVALIS Inst Vegetal, Serv Traitement Informat & Methodol, Expt Stn, F-91720 Boigneville, France

(4) INRA, UMR Agronomie, AgroParisTech, Université Paris-Saclay, 78850 Thiverval Grignon, France

2.1 Abstract

Septoria leaf blotch is a widespread disease caused by the fungus *Zymoseptoria tritici* (formerly known as *Mycosphaerella graminicola*). It causes yield losses in winter wheat crops (*Triticum aestivum* L.) in many European countries. In this study, we aimed to develop statistical models for estimating regional and site-specific incidence of *Septoria* leaf blotch from in-season field measurements. Four generalised linear models and four generalised linear mixed-effect models were fitted to six years of data collected from a major wheat-producing area of France, using frequentist and Bayesian methods. We compared the abilities of these models to predict *S. tritici* incidence over different time scales. We found that the best models were those that included site-year effects and disease risk ratings based on sowing dates and cultivar resistance levels. These models can be used to estimate the dynamics of disease incidence from observations collected in regional surveys and, as such, could help regional extension services evaluate current disease incidence at the regional scale. The proposed models could also be adjusted to make use of site-specific in-season field measurements for the estimation of site-specific disease incidence. With the current survey design, site-specific estimates are more accurate than regional estimates after mid-May. Such estimates could be used to help farmers adapt their control strategies locally during the growing season.

Keywords : *Septoria tritici*; Disease management; Incidence estimation; Generalised mixed-effect linear model; Bayesian

2.2 Introduction

Septoria leaf blotch (SLB) is a widespread disease (King et al., 1983; Eyal et al., 1987; Ponomarenko et al., 2011) causing yield losses in many European countries (Tabib Ghaffary et al., 2011; Jørgensen et al., 2014; Fones and Gurr 2015), including France (El Chartouni et al., 2012). SLB is caused by the fungus *Zymoseptoria tritici* (formerly known as *Mycosphaerella graminicola*) (Fuckel) Schröter. SLB depends on temperature (mostly during inoculation), rainfall, and wind (Pietravalle et al., 2003; Henze et al., 2007; Te Beest et al., 2009), but also on plant architecture (Robert et al., 2008). Young wheat plants are infected during autumn, by ascospores on plant residues (Eyal et al., 1987), dispersed by wind (Suffert et al., 2011). The disease spreads from leaf to leaf in spring. Pycnidia develop on leaves, releasing pycnidiospores,

which spread the disease to the upper leaves and to nearby plants, mostly by splash (King et al., 1983; Eyal et al., 1987; Tabib Ghaffary et al., 2011; Ponomarenko et al., 2011).

Several authors have suggested that the intensification of agriculture (including increase in sowing density, N fertilizer doses and the frequency of wheat in crop rotations in particular) has led to a build-up of SLB inoculum in Europe (Ansar et al., 2010; Ponomarenko et al., 2011; El Chartouni et al., 2012). Despite the use of resistant cultivars and of other non-chemical management methods for reducing the risk of SLB (e.g., use of non-host crops in crop rotations) (Eyal et al., 1987; Ponomarenko et al., 2011), fungicides are frequently applied to control the disease (two or three fungicides per year, Arvalis - Institut du végétal 2015). The incidence of SLB varies between years and sites, but many farmers systematically apply fungicides to prevent SLB, even in crop fields in which SLB incidence and severity are not sufficiently high to cause significant yield loss (Verreet et al., 2000; Jørgensen et al., 2014).

Warning systems have been developed in several countries, including the USA (Sim et al., 1998; Sine et al., 2010), UK (Polley and Thomas 1991; Hardwick et al., 2001; Parsons and Te Beest 2004; Sine et al., 2010; Jørgensen et al., 2014), Denmark, Germany, The Netherlands, Sweden and Hungary (Jørgensen et al., 2014). In 2008, the EcoPhyto plan was implemented in France, as a means of decreasing pesticide use. The main objective of this plan is to detect, identify and track emerging plant health risks (MAAF 2008). In order to reach this objective, an alert system based on yearly regional disease surveys was developed in France (Regnault-Roger 2014). Each regional survey monitors the incidence (proportion of diseased plants or organs) and/or severity (proportion of plant tissue that is symptomatic) of major crop pests and diseases in several selected plots from the region. Regional disease surveys became mandatory in 2009 (MAAF 2008; MAAF 2015a). The incidences and/or severities measured in regional surveys are published in a regional plant health bulletin (« Bulletin de Santé du Végétal ») distributed free of charge. This plant health bulletin provides no explicit recommendations concerning management, but its contents should help farmers and their advisors to decide whether there is a need to apply chemical treatments (MAAF 2008).

Champagne-Ardenne is a major wheat-producing region in France. The risk of SLB in this region is very high, but varies considerably between years (Decarrier and Gagliardi 2015). SLB incidence in this region is monitored in surveys of about 70 fields each year, chosen by regional agricultural extension services. In each of the selected fields, 20 plants per week are sampled

during the spring, and the last three leaves of each sampled plant are observed to determine the presence or absence of SLB. The observations are reported weekly in the plant health bulletin for Champagne-Ardenne. These data are potentially very useful, but are not subjected to any statistical analysis. Only simple summary statistics (mean, minimum, and maximum incidences) are published in the plant health bulletin and interpreted so as to provide general information. Site-specific and regional disease incidence is not estimated, and confidence intervals are not calculated from the survey data. No statistical model is currently used to estimate SLB incidence in real time during the wheat-growing season.

According to Van Maanen and Xu (2003), models can serve several purposes such as predicting disease infection, estimating the frequency of epidemics, and comparing the performance of different disease management strategies. Models can be incorporated in decision support systems to manage crop protection (Rossi et al., 2012; Small et al., 2015). Several process-based dynamic models have been developed for SLB; PROCULTURE (Moreau and Maraite 1999; El Jarroudi et al., 2009), EPIWHEAT (Savary et al., 2015), PRESEPT (Regnault-Roger 2005), Septo3D (Robert et al., 2008; Penaud et al., 2011) and SeptoLIS (Gouache et al., 2012) based on the model of Audsley et al., (2005). These models are useful tools for simulating disease dynamics as a function of local weather conditions, cultivar characteristics, and farmers practices. They include however, a high number of input variables and can only be run on a limited number of sites. For example, in the Champagne-Ardenne region, SeptoLIS is currently run on only four sites by the regional extension service (CRACA 2012). Although process-models provide useful information to farmers, they cannot be easily used to summarise regional survey data collected by extension services on a large number of sites. Statistical models have an important role to play because they can be easily fitted to regional survey incidence data with updating as new data become available (weekly in the case of SLB in Champagne-Ardenne). Statistical models require that a set of data be collected in sites where an estimate of disease incidence is to be made. As regional disease surveys became mandatory in 2009, observations of SLB incidence are now collected yearly on a large number of sites in major French wheat-producing regions. As survey data are available, it is possible to use statistical models for estimating regional disease incidence and analysing disease incidence variability across sites and years.

The objective of this paper was to develop statistical models for estimating regional and site-specific SLB incidence from the regional survey data collected sequentially during the growing

season. The purpose of this paper is not to produce new insights into the epidemiology of SLB, but to present statistical models for estimating SLB dynamics from regional surveys.

Several generalised linear models and generalised linear mixed-effect models were fitted to six years of data collected in Champagne-Ardenne, using frequentist and Bayesian methods. Some of the models tested included risk ratings as explanatory variables. We compared the models, using several criteria, and analysed the uncertainty associated with model estimates. Our models take into account soil and climate variables influencing *S. tritici* through the in-season field disease incidence measurements used to estimate the model parameters. We show how these models add value to plant health bulletins and help farmers and their advisors analyse risk of SLB.

2.3 Materials and methods

2.3.1 Data

The data were extracted from the Vigicultures® database (Sine et al., 2010). This database includes observations from surveys in France since 2009 (Simonneau et al., 2010; Sine et al., 2010). The surveys were performed to support an alert system developed in the framework of the EcoPhyto plan implemented as a means of decreasing pesticide use. We used the data collected for *S. tritici* in Champagne-Ardenne, one of the major wheat-producing regions in France (DRAAF-CA 2014). Depending on the number of observers, between 65 and 94 fields (mostly farmers' fields) were selected annually for disease monitoring, from 2009 to 2014. The fields were chosen before the start of the growing season and were distributed over the whole region. Climate was not monitored at the field scale and disease severity was not measured. *S. tritici* incidence was measured according to the protocol of Simonneau et al., (2011b), weekly, from March to June in 18 by 20 m plot left untreated in each selected field. At each date of observation, incidence was measured for the last three leaves, from samples of 20 wheat plants selected at random from each plot. Each leaf was visually inspected for the presence of SLB. Leaf numbers were corrected to take crop development after the date of observation into account. In accordance with the protocol of Simonneau et al., (2011b), wheat sowing date and wheat cultivar were reported for each field monitored. The total number of site-years included in the dataset was 421, resulting in 6793 observations of incidence (**Table 2.1**).

Table 2.1: Characteristics of disease incidence data for leaves 1, 2 and 3.

Leaf	Observed incidence values (%)									Number of site-years with low risk level	Number of site-years with medium risk level	Number of site-years with high risk level	Total number of site-years	Total number of observations
	April			May			June							
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max					
Leaf 1	0	0	0	0	0.80	100	0	23.98	100	92	196	115	403	1948
Leaf 2	0	2.12	70	0	5.53	100	0	43.41	100	93	198	117	408	2238
Leaf 3	0	2.31	100	0	19.47	100	0	61.54	100	97	205	119	421	2607

The correspondence between final leaf numbers (L1, L2, L3; L1 being the uppermost leaf) and observed leaf numbers is presented in **Annexe 1**. Different fields were included in the survey each year, generating a variable corresponding to a unique combination of field and year, referred to hereafter as « site-year ». Depending on the field and leaf number, there were one to 18 observation dates per site-year, with a mean of nine. As an example, the disease incidence data collected at two sites in 2013 for the leaves L1, L2, L3 are shown in **Fig. 2.1**. Each selected field is characterized by a qualitative SLB risk rating (low, medium, or high risk). This rating was determined from wheat sowing date and wheat cultivar resistance, on the basis of a simple classification rule defined by the Champagne-Ardenne extension service (**Table 2.2**).

Table 2.2: Definition of the three risk ratings used to categorise site-years

Risk rating	Sowing date	Cultivar resistance to SLB
Low	Any date	6.5-7
Low	After 24/10	5-5.5-6
Medium	After 24/10	4-4.5
Medium	Before 25/10	5-5.5-6
High	Before 25/10	4-4.5

Cultivar resistance is the ability of the cultivar to avoid developing the disease, with highly resistant varieties having a low risk of disease, whereas less resistant varieties have a higher risk of disease (GEVES 2014).

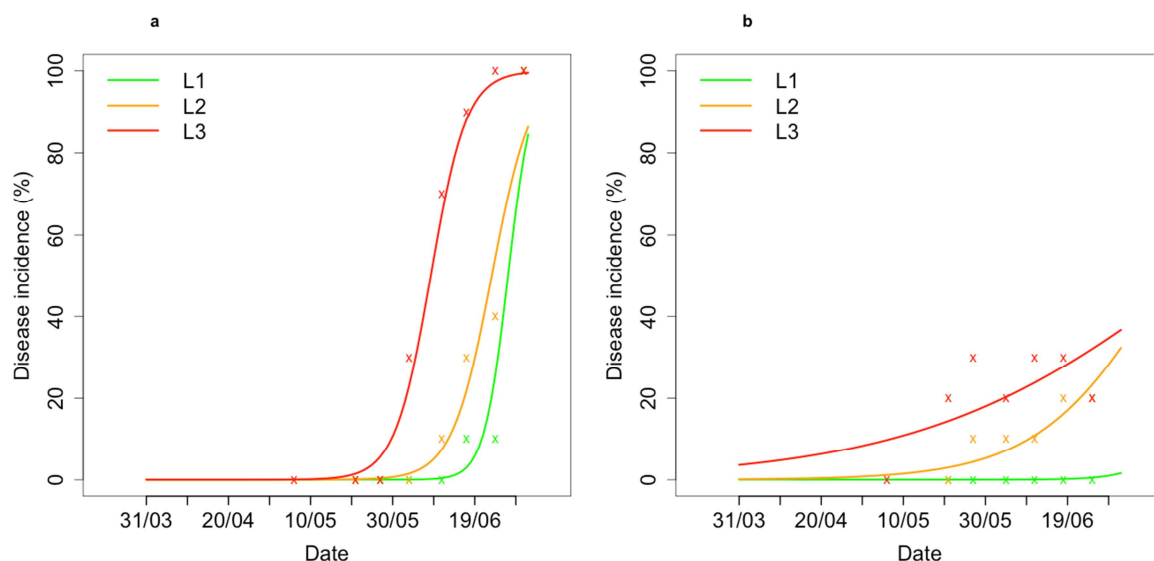


Figure 2.1: Disease incidence data (%) collected in 2013 (points) for the three final wheat leaves (L1,L2,L3) at one site with high risk (a) and one site with low risk (b). Site-specific estimates obtained with model 4F (solid lines) are also presented

2.3.2 Statistical models

In this study, we considered both frequentist and Bayesian statistical models. In the frequentist approach, the model parameters are fixed, but the parameter estimators are random, because they depend on observations. In the Bayesian approach, the parameters are defined as random variables and prior and posterior probability distributions are used to represent beliefs about parameter values before and after observation. Posterior probability distribution is derived from the prior distribution and a sample of data. It can be used to calculate credibility intervals describing uncertainty in model outputs. The introduction of iterative computational techniques (i.e., Markov Chain Monte Carlo) has increased the use of Bayesian methods in agricultural sciences (Wallach et al., 2014), but the relative usefulness of Bayesian and frequentist models has been discussed in only a few studies in the domain of plant pathology (Mila and Carriquiry 2004; Mila and Ngugi 2011; Makowski et al., 2014).

Here, eight generalised linear models (four frequentist models and four Bayesian models) were compared for estimating SLB incidence as a function of time (**Table 2.3**).

Table 2.3: Criteria used to evaluate the models: Akaike information criterion (AIC), deviation information criterion (DIC), and root mean square error (RMSE). Two types of RMSE were calculated for models including random site-year effects (2F, 4F, 2B, 4B): the RMSE for median estimates and the RMSE for site-year-specific estimates (in brackets). AIC, DIC and RMSE were calculated separately for the three types of leaves.

Model	Statistical method	Risk level effect	Site-year random effect	Leaf 1			Leaf 2			Leaf 3		
				AIC	DIC	RMSE	AIC	DIC	RMSE	AIC	DIC	RMSE
1F	Frequentist	No	No	14274		22.56	19994		24.63	27417		27.23
2F	Frequentist	No	Yes	4279		24.76 (5.37)	7360		26.57 (7.30)	11547		28.61 (10.37)
3F	Frequentist	Yes	No	13727		22.12	19420		24.21	26914		26.90
4F	Frequentist	Yes	Yes	4261		24.33 (5.37)	7329		26.13 (7.30)	11532		28.23 (10.37)
1B	Bayesian	No	No		14274	22.56		19994	24.63		27417	27.23
2B	Bayesian	No	Yes		3222	24.65 (5.44)		6039	26.54 (7.40)		9801	28.56 (10.52)
3B	Bayesian	Yes	No		13727	22.12		19420	24.21		26914	26.90
4B	Bayesian	Yes	Yes		3219	24.32 (5.44)		6033	26.10 (7.40)		9802	29.18 (10.52)

Model 1F is a fixed-parameter model of the probability of a leaf being diseased as a function of time. This model includes only two parameters and does not account for the between-site-year variability of SLB incidence or the risk ratings of plots. Model 2F is a mixed-effect model including random terms describing the between-site-year variability of disease incidence. Model 3F is a fixed parameter model including risk rating as an explanatory variable (in addition to time). Model 4F is a mixed-effect model taking both between-site-year variability and risk rating. Models 1F-4F are frequentist and were fitted to the data by maximum likelihood methods. Models 1B-4B are similar to models 1F-4F, but fitting was achieved using a Bayesian method. All the models were fitted separately to the data collected for leaves L1, L2, and L3. The model is presented below.

In all models, the number of infected leaves was assumed to follow a binomial probability distribution defined by $Y_{ij} \sim \text{Binomial}(n, \pi_{ij})$, where Y_{ij} is the number of infected leaves at the j^{th} date of observation, $j=1, \dots, M_i$, among $n=20$ leaves sampled in the i^{th} site-year, $i=1, \dots, N$ (N was equal to 403, 408, 421 for leaves L1, L2 and L3, respectively) and π_{ij} is the probability of a leaf being diseased at the j^{th} date of observation in the i^{th} site-year.

Model 1F is defined as follows: $\text{logit}(\pi_{ij}) = \alpha_0 + \alpha_1 \times t_{ij}$, where t_{ij} is the date of the j^{th} observation in the i^{th} site-year, and α_0 and α_1 are two fixed parameters.

Model 2F is a random-effect model defined as follows: $\text{logit}(\pi_{ij}) = \alpha_0 + b_{0i} + (\alpha_1 + b_{1i}) \times t_{ij}$, where α_0 and α_1 are two fixed parameters, and b_{0i} and b_{1i} are two random parameters distributed as $\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} \sigma_{b_0}^2 & c \\ c & \sigma_{b_1}^2 \end{bmatrix}$, $\sigma_{b_0}^2$ and $\sigma_{b_1}^2$ are the variances of the random parameters describing the extent of the between-site-year variability of SLB incidence, and c is the covariance between the two random terms.

Model 3F is a fixed parameter model including a covariable describing the level of SLB risk. It is defined as follows: $\text{logit}(\pi_{ij}) = \alpha_0 + g_M \times Z_{Mi} + g_H \times Z_{Hi} + \alpha_1 \times t_{ij}$, where Z_{Mi} and Z_{Hi} are two binary variables ($Z_{Mi} = 1$ indicates that the i^{th} field is characterized by a risk rating of “medium”, with $Z_{Mi} = 0$ otherwise, $Z_{Hi} = 1$ indicates that the i^{th} field is characterized by a risk rating of “high”, with $Z_{Hi} = 0$ otherwise), α_0 , α_1 , g_M , and g_H are fixed parameters.

Model 4F is a random-effect model including a covariable describing the level of SLB risk. It is defined as follows: $\text{logit}(\pi_{ij}) = \alpha_0 + b_{0i} + g_M \times Z_{Mi} + g_H \times Z_{Hi} + (\alpha_1 + b_{1i}) \times t_{ij}$, where α_0 , α_1 , b_{0i} , b_{1i} , g_M , and g_H are defined as above.

The parameters of models 1F-4F are estimated using the maximum likelihood method, implemented with the R function `glm` for models 1F and 3F (R Development Core Team 2014), and function `glmer` for models 2F and 4F (package `lme4` version 1.1-6). The R script used to fit model 4F is presented in **Annexe 2**.

The Bayesian models, 1B-4B, are based on the same equations as models 1F-4F, but their parameters are estimated by a Bayesian approach. For models 1B-4B, the following weakly informative (with large variances) prior distributions were used; $\alpha_0, \alpha_1, g_M, g_H \sim N(0, 10^6)$, $\Sigma \sim \text{Wishart}(\psi, \nu)$, with $\psi = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$, and $\nu = 2$.

For the Bayesian models including site-year effect (2B and 4B), it was not possible to achieve convergence for leaf 1 when the parameters b_{0i} and b_{1i} were assumed to be correlated. These two models were thus fitted with uncorrelated parameters.

Posterior distributions of the parameters of the models were estimated with winBUGS software (Lunn et al., 2000), using a Markov chain-Monte Carlo algorithm with 100,000 iterations and a burn-in period of 50,000 iterations. Convergence was checked by the Gelman-Rubin method

(Gelman and Rubin 1992) implemented with three chains of parameter values. The winBUGS script used to fit model 4B is presented in **Annexe 3**.

2.3.3 Model assessment

Model residuals were analysed graphically. We calculated the Akaike information criterion (AIC) (Akaike 1974) for models 1F-4F, and the deviance information criterion (DIC) (Spiegelhalter et al., 2002) for models 1B-4B. The best models are those with the lowest AIC or DIC values.

The accuracy of model outputs was evaluated by calculating the root mean square error (RMSE) and the root mean square error of prediction (RMSEP). With model 1F, incidence at time t_{ij} is predicted by

$$\pi_{ij} = \frac{\exp(\alpha_0 + \alpha_1 \times t_{ij})}{1 + \exp(\alpha_0 + \alpha_1 \times t_{ij})} \quad (1)$$

and, with model 3F, SLB incidence at time t_{ij} is predicted by

$$\pi_{ij} = \frac{\exp(\alpha_0 + g_M \times Z_{Mi} + g_H \times Z_{Hi} + \alpha_1 \times t_{ij})}{1 + \exp(\alpha_0 + g_M \times Z_{Mi} + g_H \times Z_{Hi} + \alpha_1 \times t_{ij})} \quad (2)$$

where parameter values are set at their maximum likelihood estimates. With models 1B and 3B, SLB incidence is predicted by the posterior means of π_{ij} (Eqs. (1,2)) calculated from the chains of parameter values generated. Two types of estimates were established for the four models including site-year effects (2F, 2B, 4F, 4B): median and site-specific estimates. Median estimates were obtained with Eqs. (1,2), as explained above. Site-specific estimates were derived from the estimated values of the site-year effects b_{0i} and b_{1i} . With model 2F, the estimate for the i^{th} site-year at time t_{ij} is thus expressed as

$$\pi_{ij} = \frac{\exp(\alpha_0 + b_{0i} + (\alpha_1 + b_{1i}) \times t_{ij})}{1 + \exp(\alpha_0 + b_{0i} + (\alpha_1 + b_{1i}) \times t_{ij})} \quad (3)$$

where α_0 and α_1 are equal to their maximum likelihood estimates, and b_{0i} and b_{1i} are equal to their best linear unbiased predictors (BLUP). Similarly, with model 4F, the estimate for the i^{th} site-year at time t_{ij} is expressed as

$$\pi_{ij} = \frac{\exp(\alpha_0 + b_{0i} + g_M \times Z_{Mi} + g_H \times Z_{Hi} + (\alpha_1 + b_{1i}) \times t_{ij})}{1 + \exp(\alpha_0 + b_{0i} + g_M \times Z_{Mi} + g_H \times Z_{Hi} + (\alpha_1 + b_{1i}) \times t_{ij})} \quad (4)$$

where α_0 , α_1 , g_M , g_H are taken as their maximum likelihood estimates, and b_{0i} and b_{1i} are equal to their BLUP. With models 2B and 4B, site-specific incidence is predicted by the posterior means of π_{ij} (Eqs. (3,4)) calculated from the chains of parameter values generated.

For each type of estimation and each leaf (L1, L2, L3), RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{M} \sum_{i,j} (O_{ij} - P_{ij})^2} \quad (5)$$

where O_{ij} and P_{ij} are the observed and estimated incidence in site-year i at time t_{ij} , and M is the total number of data ($M=1948$ for L1, $M=2238$ for L2, and $M=2607$ for L3). For models 2F, 4F, 2B, and 4B, separate RMSE values were calculated for median estimates and for site-specific estimates. In Eq. (5), SLB incidence estimates are made for all data, using the same dataset for both parameter and model estimation. This approach may lead to an underestimation of model prediction errors (Wallach et al., 2014). We dealt with this issue by also evaluating model predictions by calculating RMSEP (Wallach et al., 2014). RMSEP is based on Eq. (5), but is calculated from independent incidence data rather than from the data used for parameter estimation. Specifically, four prediction dates were considered successively: May 1st, May 15th, May 30th, and June 15th. SLB incidence was predicted each year for the three weeks following each prediction date, for all sites, without the use of the data collected after the prediction date. The performance of the model was evaluated on data not included in the sample used to estimate the model parameters. A separate RMSEP was then calculated for each prediction date, model and leaf type. This approach made it possible to obtain out-of-sample forecast performances for the statistical models and to assess the predictive capability of the models considered at different prediction dates. For example, **Figure 2.2** shows examples of site-specific predictions obtained with model 4F from May 15th to June 4th for three sites in 2014. The data available between May 15th and June 4th (closed symbols in **Figure 2.2**) were used to evaluate the prediction errors but not to estimate the model parameters.

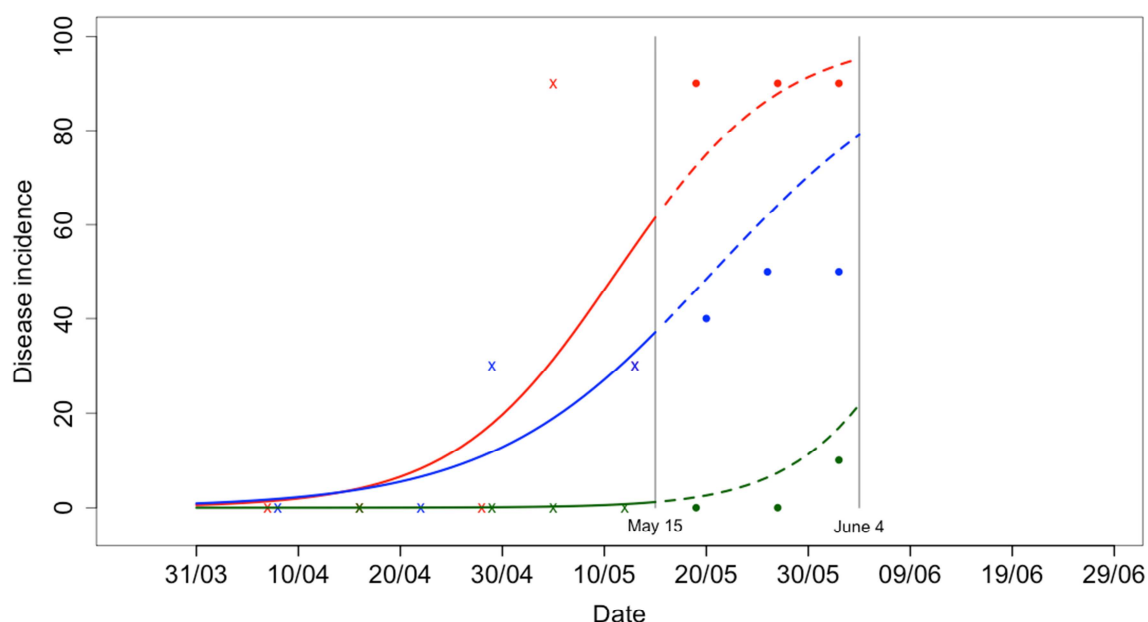


Figure 2.2 : Disease incidence data (%) and site-specific estimates obtained with model 4F for three different sites in 2014 for leaf 3. Site-specific estimates obtained before May 15th and after May 15th are indicated by solid and dashed lines respectively. Observations collected after May 15th (closed symbols) were not used to fit the model

2.4 Results

2.4.1 Estimated parameter values

The estimated parameter values obtained with the frequentist method were very similar to the values obtained with the Bayesian method (**Table 2.4** and **Annexes 4** and **5**). For example, the estimated value of α_1 for model 4F was 0.168 for leaf 2, a value very close to the posterior mean value estimated for model 4B for the same leaf (0.170) (**Table 2.4**). The 95 % confidence intervals calculated with the frequentist method and the 95 % credibility intervals computed with the Bayesian technique were also similar (**Table 2.4** and **Annexes 4** and **5**).

Table 2.4: Estimated parameter values / estimated posterior means of parameters and 95% confidence intervals / credibility intervals (in brackets) obtained for the frequentist model 4F / Bayesian model 4B.

Parameters	Leaf 1	
	Model 4F	Model 4B
α_0	-16.971 (-18.503 ; -15.438)	-17.230 (-18.57 ; -16.04)
α_1	0.238 (0.214 ; 0.263)	0.243 (0.224 ; 0.263)
g_H	2.550 (1.439 ; 3.661)	2.649 (1.624 ; 3.696)
g_M	0.942 (-0.096 ; 1.980)	1.005 (0.048 ; 2.014)

Parameters	Leaf 2	
	Model 4F	Model 4B
α_0	-11.641 (-12.472 ; -10.811)	-11.717 (-12.5 ; -10.88)
α_1	0.168 (0.155 ; 0.181)	0.17 (0.156 ; 0.183)
g_H	1.766 (1.072 ; 2.459)	1.804 (1.135 ; 2.517)
g_M	0.250 (-0.394 ; 0.894)	0.27 (-0.375 ; 0.942)

Parameters	Leaf 3	
	Model 4F	Model 4B
α_0	-9.519 (-10.176 ; -8.862)	-9.389 (-10.15 ; -8.688)
α_1	0.133 (0.123 ; 0.144)	0.131 (0.118 ; 0.143)
g_H	1.195 (0.648 ; 1.743)	1.231 (0.753 ; 1.782)
g_M	0.509 (0.011 ; 1.007)	0.533 (0.03 ; 1.057)

The inclusion of random site-year effects had a large effect on the estimated parameter values. For example, for leaf 1, the estimated value of α_0 was -7.223 for model 1F and -15.854 for model 2F (**Annexes 4** and **5**). Confidence and credibility intervals were larger for models including random site effects than for models not including site-year effects, particularly for α_0 but also for the other parameters, albeit to a lesser extent (**Table 2.4** and **Annexes 4** and **5**). The larger confidence intervals resulted from the models including random site-year effects assuming that data collected in the same site-year were not independent. These models thus make more realistic assumptions about the distribution of the data.

The sizes of confidence and credibility intervals tended to decrease from leaf 1 to leaf 3. For example, in model 4F, the difference between the upper and lower limits of the confidence interval for the estimated value of α_1 was 0.049 for leaf 1, 0.026 for leaf 2, and 0.021 for leaf 3.

Similar trends were observed for the other parameters, because the amount of data increased from leaf 1 to leaf 3, with 1948 observations for leaf 1, 2238 for leaf 2, and 2607 for leaf 3. Leaf 1 emerged later in the season than leaves 2 and 3, resulting in the collection of fewer data for this leaf. Parameter values were therefore less accurately estimated for leaves 1 and 2 than for leaf 3.

2.4.2 Model selection and quality of fit

Models including a random site-year effect had much lower AIC and DIC values (**Table 2.3**). Model 2F had a lower AIC value than model 1F, and model 2B had a lower DIC value than model 1B. Based on these criteria, models 2F and 2B thus performed better than models 1F and 1B. The inclusion of the risk rating variables Z_{Mi} and Z_{Hi} had only a small effect on the AIC and DIC (**Table 2.3**). The values of these criteria were lower when risk rating variables were included, but the differences were small relative to models not including these variables.

RMSE values increased from leaf 1 to leaf 3. For median frequentist estimates, they ranged from 22.12 to 24.76 for leaf 1, from 24.21 to 26.57 for leaf 2, and from 26.90 to 28.61 for leaf 3 (**Table 2.3**), reflecting the increase in incidence from leaf 1 to leaf 3 (**Table 2.1**). Models including risk-rating variables tended to have smaller RMSE values than simpler models not including such variables, but the differences were small. For example, for leaf 1 and median estimates, the RMSE of model 1F (without risk ratings) was 22.56, whereas that of model 3F (with risk ratings) was 22.12 (**Table 2.3**). Site-year-specific estimates yielded a lower RMSE than median estimates. RMSE values differed by more than 60 %, revealing the strong between-site-year variability of the dynamics of disease incidence. Model estimates were closer to the observed values when site-year effects were taken into account. Almost no difference in RMSE was found between frequentist and Bayesian models (**Table 2.3**).

The better quality of fit for site-year-specific estimates was confirmed by the analysis of model residuals (**Fig. 2.3**). The residuals of models not including site-year effects (**Figs. 2.3a, c**) were distributed asymmetrically around zero. The distribution of these residuals is skewed to the right in April-May, and skewed to the left in June-July, i.e., at the end of the growing season (**Figs. 2.3a, c**). Distributions of residuals were more symmetric for models including site-year effects (**Figs. 2.3b, d**). Similar distributions were obtained for the Bayesian models and for the other two leaves (not shown). The differences between models with and without site-year effects

were due to the considerable site-year variability of disease dynamics. In models including no site-year effect, a single disease incidence dynamic was simulated for all site-years. In site-years in which the disease developed rapidly, model simulations underestimated disease incidences, generating positive residuals. Conversely, in site-years in which the disease developed much more slowly, model simulations overestimated disease incidences, generating negative residuals. Models including site-year effects more accurately reproduced the between-site-year variability of the disease dynamics, resulting in a much more symmetric distribution of residuals.

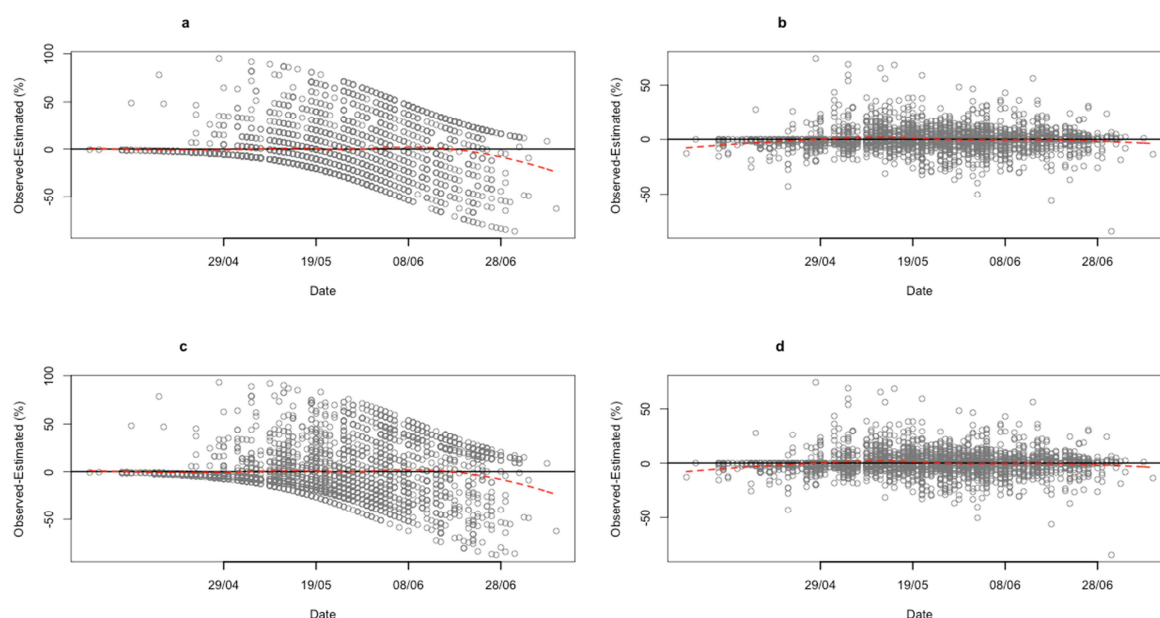


Figure 2.3 : Residuals (observed incidence – estimated incidence) obtained with the frequentist models for leaf 2 as a function of time. The dashed red curves were obtained by local polynomial regression (loess). a. Residuals of model 1F. b. Site-year-specific residuals of model 2F. c. Residuals of model 3F. d. Site-year-specific residuals of model 4F

2.4.3 Model predictions

No substantial difference in RMSEP values was found between the frequentist and Bayesian models. We therefore present the RMSEP values for the frequentist models only.

RMSEP values increased during the growing season, with smaller prediction errors for May 1st - 15th than for May 30th - June 15th (**Figs. 2.4** and **2.5**). This reflects the relatively low disease incidences observed in early May, with a substantial proportion of these observations close to zero, particularly for leaf 1 (**Table 2.1**, **Fig. 2.1**). Later in May and in June, the variability of the observed incidences was much higher and the prediction errors tended to be larger.

In most cases, the RMSEP values obtained two and three weeks after the prediction dates were higher than those obtained one week after the prediction date. Thus, long-term predictions are generally less accurate than short-term predictions. The only exceptions were the RMSEP values obtained for model 1F and model 3F on June 15th for leaf 1 and on both May 30th and June 15th for leaf 3. After these dates, no major change in disease incidence was observed over time, with similar RMSEP values obtained for predictions made one, two or three weeks in advance.

RMSEP values increased from leaf 1 to leaf 3, for the first three prediction dates (May 1st, May 15th, and May 30th). For example, with model 1F, the RMSEP values for predictions obtained one week ahead on May 1st were 1.2% for leaf1 and 17.1% for leaf 3.

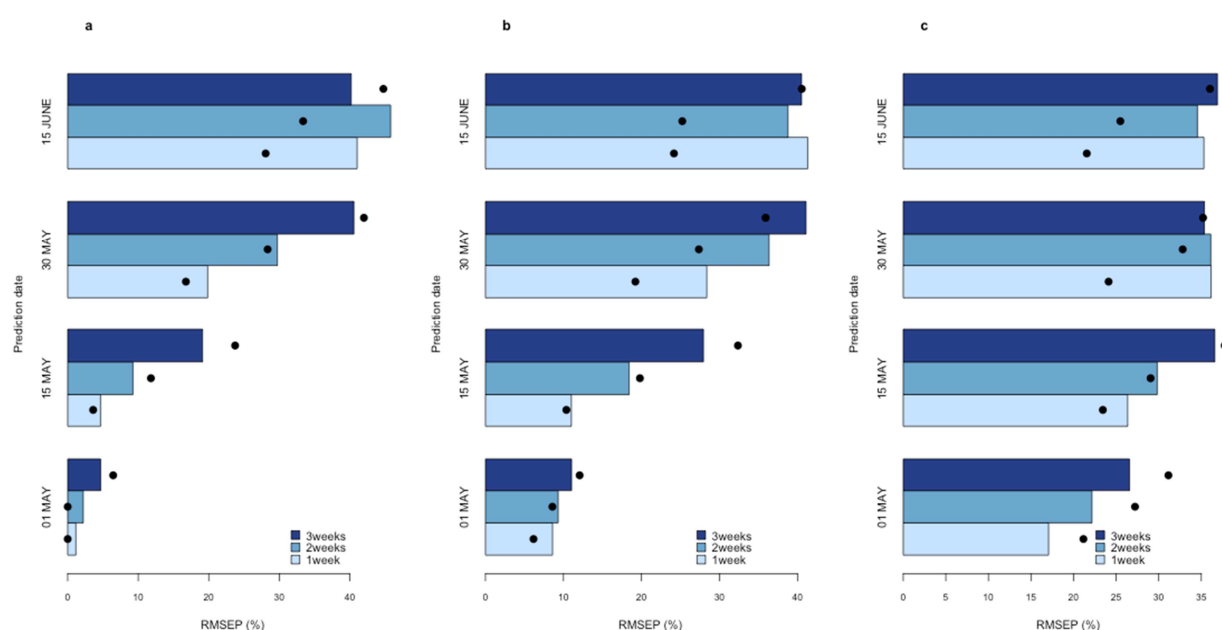


Figure 2.4 : Root mean square error of prediction (RMSEP %) for four different prediction dates (May 1st, May 15th, May 30th, June 15th) and three time lags (predictions made 1 week, 2 weeks, 3 weeks in advance). Bars indicate the RMSEP of model 1F and points indicate the RMSEP of model 2F (site-year-specific predictions). a. RMSEP for leaf 1. b. RMSEP for leaf 2. c. RMSEP for leaf 3

This reflects the lower incidence and its variability for leaf 1 than for leaves 2 and 3 in May (**Table 2.1**). Prediction errors thus tended to be smaller for leaf 1 than for leaves 2 and 3.

After May 15th, most of the RMSEP values were smaller for the site-year-specific predictions of models 2F and 4F than for the median predictions of models 1F and 3F. For May 30th and June 15th, disease incidences tended to be predicted most accurately by site-year-specific predictions. Before these dates, there were too few site-year-specific measurements for site-year-specific predictions to be clearly advantageous, and accuracy was similar for the two types of predictions. Moreover, although site-specific predictions tend to be more accurate than median predictions for short-term predictions, the differences in RMSEP were smaller for long-term predictions (i.e., three weeks ahead).

The RMSEP values obtained on May 1st were smaller for the models including risk rating variables (3F and 4F, **Fig. 2.5**) than for models without risk rating variables (1F and 2F, **Fig. 2.4**). Early predictions were thus more accurate if calculated taking risk variables into account. For example, for one-week-ahead predictions obtained on May 1st for leaf 2 with model 2F the RMSEP was 6.16 %, whereas, for similar predictions with model 4F (including risk rating variables) the RMSEP was only 3.34 %. These differences became negligible after May 1st.

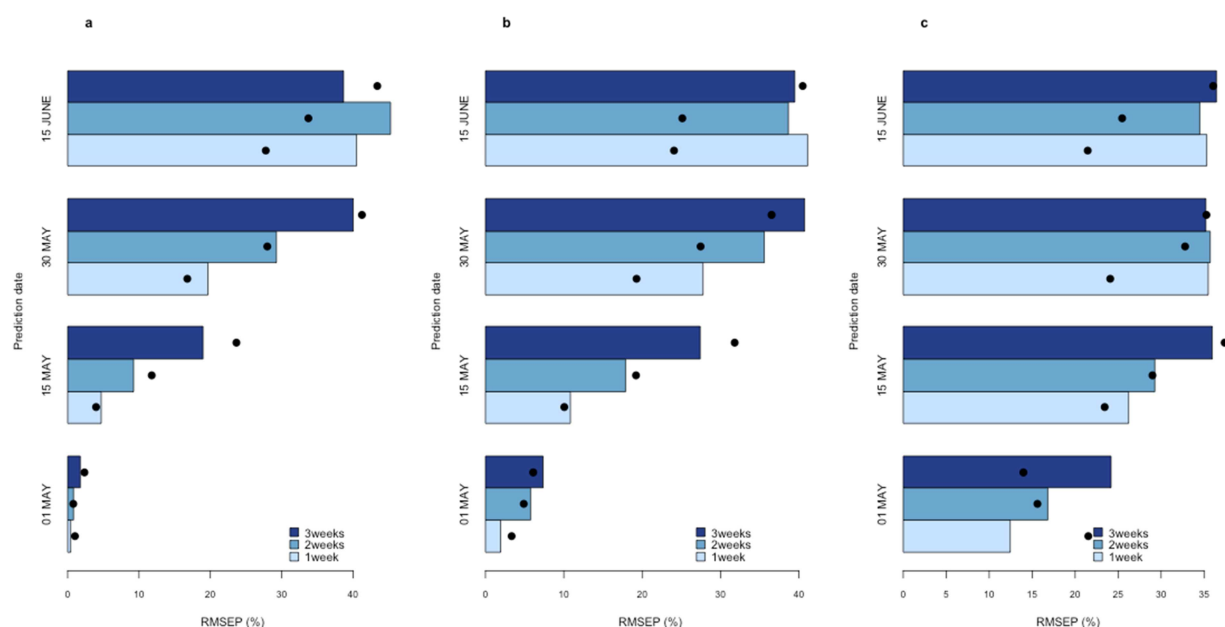


Figure 2.5: Root mean square error of prediction (RMSEP %) for four different prediction dates (May 1st, May 15th, May 30th, June 15th) and three time lags (1 week, 2 weeks, 3 weeks in advance). Bars indicate the RMSEP of model 3F and points indicate the RMSEP of model 4F (site-year-specific predictions). a. RMSEP for leaf 1. b. RMSEP for leaf 2. c. RMSEP for leaf 3

2.4.4 Posterior distribution incidence

As mentioned above, frequentist and Bayesian models performed similarly for all criteria. However, one advantage of Bayesian models is that they can be used to describe the uncertainty associated with site-year-specific estimation through probability distributions and credibility intervals. This is illustrated in **Fig. 2.6**, which shows the posterior distribution obtained on May 23rd 2014 with model 4B for a given site-year. In this example, a disease incidence of 10 % is observed. The disease incidence point estimate of the Bayesian model (median of the site-specific posterior distribution) is 5.2 %, which is very close to incidence value estimated by the frequentist model 4F (4.9 %). The large 95 % credibility interval (from 0 to 70 %) derived from the posterior probability distribution indicates that a high level of uncertainty is associated with the estimated incidence value.

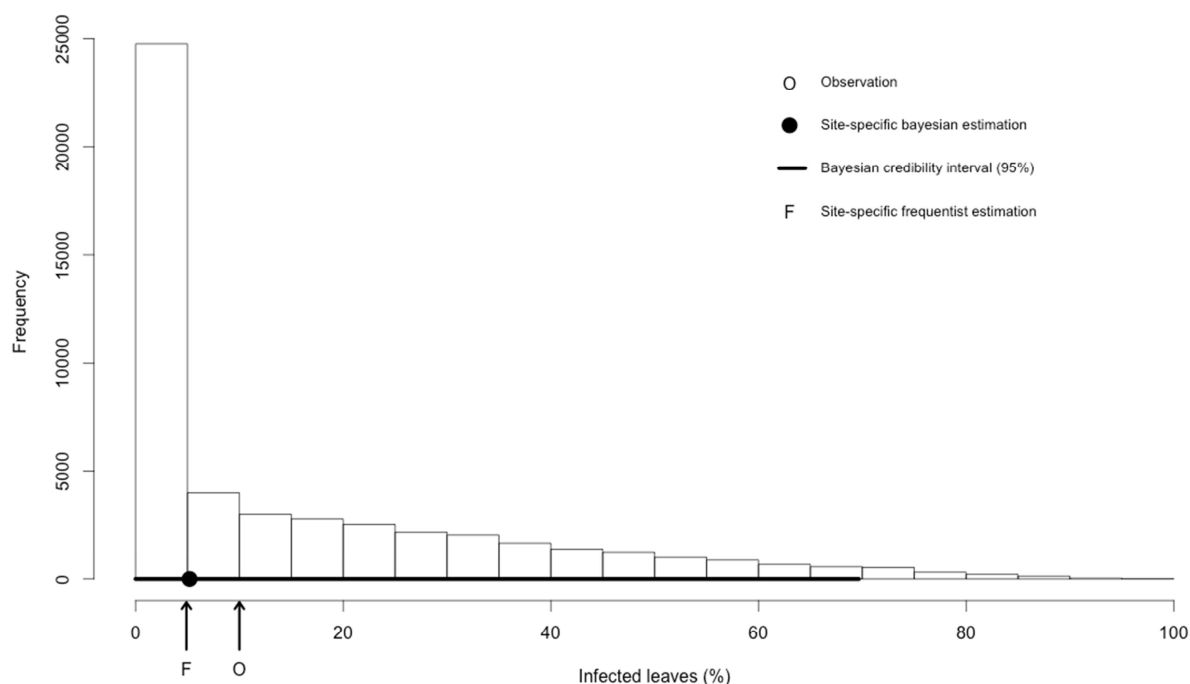


Figure 2.6: Estimated posterior distribution of disease incidence for leaf 3 on May 23rd 2014, for a site-year with an observed disease incidence of 10 %. The histogram includes 50,000 disease incidence values drawn from the posterior distribution obtained with model 4B

2.5 Discussion

Disease incidence data indicated large differences in the SLB dynamics between site-years. In some site-years, disease incidence increased very rapidly during the growing season (**Fig. 2.1a**), whereas it increased slowly in other site-years (**Fig. 2.1b**). Models including site-year effects were able to describe this variability, thereby providing a better fit, a smaller AIC/DIC (**Table 2.3**), and, in many cases, more accurate estimates. These models also yielded larger (and more realistic) confidence intervals for the estimated parameter values, because they did not consider observations collected for the same site-year to be independent (**Annexe 4** and **Table 2.4**). These findings are consistent with those of Makowski et al., (2014), who analysed the efficacy of fungicide treatments for *Phyllosticta citricarpa* and found that statistical models including site-year effects yielded larger confidence intervals. We found no substantial difference between frequentist and Bayesian models, which yielded similar parameter values and RMSEP. These results are consistent with those obtained by Mila and Ngugi (2011) and Makowski et al., (2014). Bayesian models have the advantage of being able to describe the uncertainty associated with site-year-specific estimation through a posterior probability distribution (**Fig. 2.6**). Nevertheless, these models require long computation times compared to frequentist models (for example, estimation of the parameters of models 4F and 4B required 31 s and 24 min, respectively, for leaf 3, with a computer equipped with an Intel® Core™ i7 CPU). Bayesian models could potentially take prior expert knowledge into account (independent from the available data). However, further research is required to determine what kinds of prior knowledge could be used to define prior parameter distributions. We tried to fit more complex models (including both year and site-year random effects, or including interactions between risk rating and time), but the fitting algorithm failed to converge for these models due to their complexity. Our assessments indicate that the best models are those including site-year effects and disease risk ratings (4F and 4B). A statistical model including cumulative temperature as an input variable was fitted to the data, but this model did not perform better, probably because disease incidence data for different sites partly reflect the weather conditions at those sites. Our models indirectly take into account the environmental factors influencing incidence through the in-season field disease incidence measurements used to estimate the model parameters. However, this result should not be considered general, because weather input variables have been shown to be useful for predicting disease dynamics in other situations (Gouache et al., 2012). The proposed statistical models can be used to estimate the dynamics of median disease incidence at the regional level from disease survey data. This approach is illustrated in **Fig. 2.7**.

In this example, model 4F was fitted to the disease incidence data collected before May 15th 2014. The median values were calculated across all the surveyed sites. The estimates of median disease incidence between May 15th and June 4th in 2014 (dashed brown curve in **Fig. 2.7**) are predictions, because the data collected after May 15th 2014 were not used to fit the model. Our statistical approach could be used for the quantitative analysis of disease observations collected in regional surveys monitoring SLB incidence. The results of such analyses would help farmers' advisors formulate regional recommendations for disease control management. The estimated dynamics can be updated whenever new data become available (essentially weekly in Champagne-Ardenne) and plotted in the weekly alert bulletin published by the regional extension service. The estimated dynamics of disease incidence can be compared with past dynamics, as shown in **Fig. 2.7**.

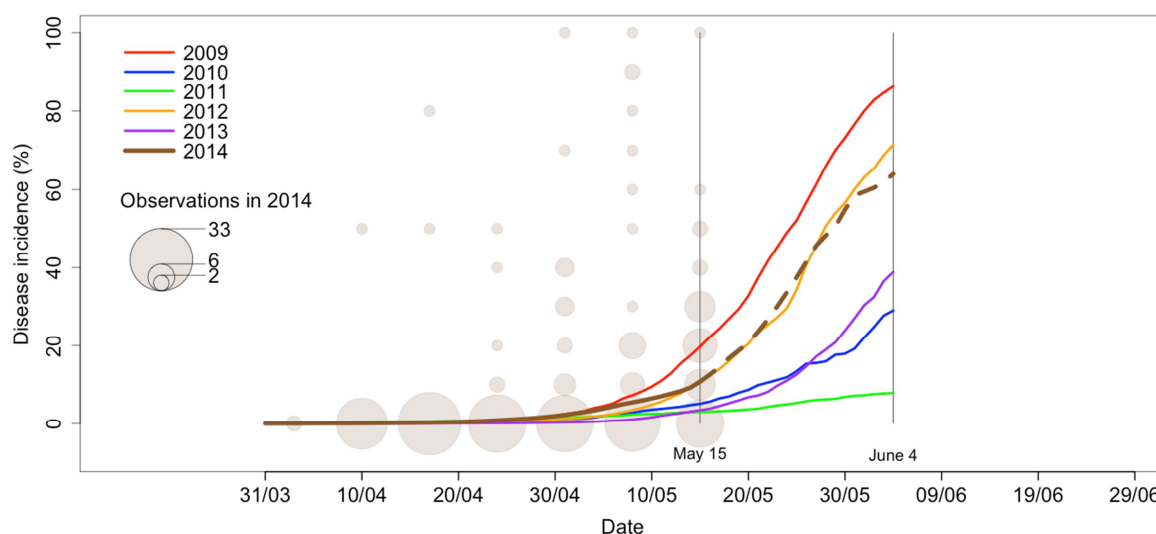


Figure 2.7 : Median regional disease incidence (%) estimated for leaf 3 from 2009 to 2014. The medians were calculated across the site- specific disease incidence dynamics estimated with the model 4F. Estimations obtained in 2014 (brown line) were based on data collected before May 15th. Estimates derived before and after May 15th are indicated by solid and dashed lines, respectively. Light brown circles indicate observed data collected in Champagne-Ardenne in 2014. Circle size is proportional to the number of data

Our models can also be used for site-specific estimations, as illustrated in **Fig. 2.2**. In this figure, model 4F was adjusted for incidence data collected before May 15th 2014, and the model was

used to predict disease incidence between May 15th and June 4th for three sites. These adjusted site-specific dynamics can be used for local diagnosis and to formulate local recommendations. However, our results indicate that it is not possible to obtain accurate site-specific predictions before mid-May when only one observation is performed per week at each site. Before this date, there are too few observations of disease incidence for accurate estimation of the site-specific dynamics of disease incidence. After this date, between one and five observations are generally available for a given site, making it possible to improve the accuracy of disease incidence predictions by adjusting the model parameters for local data. In France, fungicide treatments for SLB usually take place between Zadoks stages (Zadoks et al., 1974) Z32 and Z61 (Gouache and Couleaud 2010). In this study, these stages were reached between mid-April and June. Predictions of disease incidence obtained later than mid-May can thus help farmers adapt their control strategies after this date. An increase in the frequency of disease observations in April-May might improve the accuracy of early predictions of disease incidence.

Regional and site-specific incidence estimates could be published in a bulletin providing plant health alerts, to enable farmers and their advisors to adjust their disease control strategies several times during the course of the growing season. Such estimates would also be useful for the firms responsible for collecting the grain at harvest, as it would enable them to anticipate risks of yield loss due to high incidences of SLB.

Statistical models are no substitute for epidemiological dynamic models. Process-based dynamic models are useful for plant disease management and their outputs can be used in decision-support systems (Moreau and Maraite 2000). They can also be used to study effect of climate changes on epidemics (Barbetti et al., 2012; Kim et al., 2015; Juroszek and Von Tiedemann 2015). Several dynamic models have been developed to simulate *S. tritici* infection dynamics as a function of weather input data and farmers' practices (El Jarroudi et al., 2009; Gouache et al., 2012; Savary et al., 2015). However, the synthesis of disease incidence observations collected over the growing season by extension services is not the main objective of these dynamic models. The dynamic model SeptoLIS is thus currently run on only four sites in the Champagne-Ardenne region (CRACA 2012) and on only one site in the Centre region (CRAC 2014). Although SeptoLIS provides useful information to farmers, it cannot be easily used to summarise regional survey data collected by extension services on a large number of sites. Our statistical models can easily be fitted to regional survey incidence data. They can be updated as new data become available, and used to estimate regional disease incidence and to analyse its variability across

sites and years. We thus believe that the models presented here could make a useful contribution to alert systems developed for the monitoring of disease risks.

The statistical models presented here are suitable for analyses of the risk of *S. tritici* infection in regions in which surveys are conducted several times per year. This is the case in the main French wheat-growing areas, including Champagne-Ardenne, where a plant health bulletin containing information about plant disease risk is published and made available free of charge every week between March and July. We believe that our statistical models could improve the quality of the information displayed in this bulletin. The parameter values of the statistical models estimated for Champagne-Ardenne are valid for this region only, and should not be used in other regions. However, the model equations and estimation methods presented here define a generic framework that could easily be applied in other regions in which similar surveys are conducted for the monitoring of SLB incidence. For example, **Annexe 6** shows an application of the statistical model 4F in another major French wheat-producing region, the Centre region. This model was fitted to survey data collected in this region and used to estimate SLB dynamics from 2009 to 2015.

The codes of our models are presented in **Annexes 2** and **3**. They can be easily implemented by extension services to estimate dynamics of SLB incidence at the regional level from disease survey data. As plant disease surveys are conducted in France for several major plant diseases, it will be useful to adapt our statistical models to analyse the dynamics of other diseases, such as downy mildew caused by *Plasmopara viticola* in South-West France.

2.6 Acknowledgments

This research was supported by SynOEM - Mieux profiter de la synergie entre réseaux d'observation, expertise et modélisation pour l'élaboration du Bulletin de Santé du Végétal (<http://www.modelia.org/moodle/course/view.php?id=55>), funded by the French National Agency for Water and Aquatic Environments (ONEMA) under the Ecophyto Plan supported by the French Ministry of Agriculture. This work was cofunded by a PhD grant from the French Association for Technical Research (ANRT).

We are grateful to E. Gourdain, D. Simonneau, V. Bochu, A. Decarrier, D. Gouache, G. Hugerot, and J.Veslot for their comments on our results and to Alex Edelman & Associates for English proofreading.

Chapitre 3.

A framework based on generalised linear mixed models for analysing pest and disease surveys

Ce chapitre correspond à un article soumis le 17 février 2016 à la revue *Crop Protection* (Michel, L., Brun, F., Makowski, D., 2016. A framework based on generalised linear mixed models for analysing pest and disease surveys.).

Lucie Michel^a, François Brun^b, David Makowski^{c*}

a ACTA, INRA, UMR Agronomie, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

b ACTA, INRA, UMR AGIR, F-31326 Castanet Tolosan, France

c UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

3.1 Abstract

In many countries, regional surveys are carried out annually, to detect the presence of pests and diseases in crops. During these surveys, the incidence of major diseases and the presence of pests are recorded on various dates during the growing season. In this study, we aim to develop a framework to make better use of these regional surveys to estimate pest and disease dynamics, to analyse their variability across sites and years, and to assess uncertainty. Our framework is illustrated in four case studies: *Septoria* leaf blotch on wheat, downy mildew on grapevine, yellow sigatoka on banana and weevils on sweet potato. We used frequentist and Bayesian approaches to fit several generalised linear mixed models which can be used to estimate disease and pest dynamics from observations collected and could help regional extension services evaluate risk levels at the regional scale.

Keywords: Alert system, Bayesian model, disease survey, generalised linear mixed model, plant pest.

3.2 Introduction

Plant pest and disease alert systems provide farmers and their advisers with information about plant pest situations. They are designed to help farmers to manage their crop protection practices so as to minimise yield loss and to avoid unnecessary chemical treatments. They are usually based on regional annual surveys for major pests and diseases, which have been conducted in many countries for decades. Pest and disease surveys have been performed since 1912 in Kansas (USA (Sim IV et al., 1988)), and since the 1970s in UK (King, 1977; Polley and Thomas, 1991). Pest and disease surveys are now frequently carried out in most American and European countries (Sigvald, 2012; Sine et al., 2010; Verreet et al., 2000) and in Asian countries, including India, Indonesia, Malaysia, and the Philippines (Talukder, 2000).

In France, regional pest and disease surveys became mandatory in 2009 when the government introduced the “Ecophyto” plan to reduce pesticide uses (MAAF 2008). This plan was updated in 2015 (MAAF 2015a), and its main objective is to detect, identify and track emerging plant health risks. In France, regional surveys are now carried out annually, to detect the presence of pests and diseases in major crops. During these surveys, the incidence of major diseases and the

presence of pests are recorded on various dates during the growing season. Results are published weekly or monthly in plant health bulletins that are distributed free-of-charge to farmers. For example, in 2014, around 13000 fields were monitored in France and more than 3350 plant health bulletins were published and distributed to farmers (DGAL-SDQPV, 2014).

French plant health bulletins currently summarise survey results solely by presenting simple descriptive statistics (mean, minimum, and maximum values of the collected data) and graphics (histogram, pie chart). No statistical analysis is performed to estimate pest and disease dynamics or to compare current infestation levels with those in previous years. Another limitation of the current version of plant health bulletins is that they do not formally analyse uncertainty in the results of pest surveys.

According to Van Maanen and Xu (2003), models can serve several purposes such as predicting disease infection, estimating the frequency of epidemics, and comparing the performance of different disease management strategies. Process-based dynamic models are useful tools for simulating disease dynamics as a function of local weather conditions, cultivar characteristics, and farmers' practices (Apel et al., 2002; El Jarroudi et al., 2009; Gouache et al., 2012; Penaud et al., 2011; Whish et al., 2014). They usually include however, a high number of input variables and can only be run on a limited number of sites. For example, in the Champagne-Ardenne region (a major wheat-producing region in France), the process-based dynamic model SeptoLIS (Gouache et al., 2012) is currently run on only four sites by the regional extension service for predicting the dynamics of Septoria leaf blotch (SLB) in wheat crops (CRACA, 2012). Although process-models provide useful information to farmers, they cannot be easily used to summarise regional survey data collected by extension services on a large number of sites.

Statistical models have an important role to play because they can be easily fitted to regional survey incidence data with updating as new data become available. We propose here to make better use of regional surveys to estimate pest and disease dynamics, to analyse their variability across sites and years, and to assess uncertainty. We have developed a framework for this purpose based on generalised linear mixed models (glmm). These models are flexible enough to handle both incidence data and count data, and to deal with longitudinal measurements, such as independent measurements for the same site on different dates (Agresti, 2002 Chap.12). They can easily be fitted to data collected in pest and disease surveys, and can be updated whenever new data become available. Glmm can be used in different ways. They can be used to estimate

mean (or median) pest and disease dynamics over a population of sites and years. They can also be used to estimate dynamics from local observations for specific sites and years.

The use of glmm models has been promoted in ecology as a method for analysing nonnormal data such as counts or proportions (Bolker et al., 2009), but these models have only rarely been applied in phytopathology (Madden et al., 2002; Makowski et al., 2014; Piepho, 1999). They can be fitted to data by classical (frequentist) or Bayesian statistical methods (Fong et al., 2010; Makowski et al., 2014; Zhao et al., 2006). We demonstrate here the utility of using glmm for analysing the pest and disease survey data used in alert systems, and we compare the results obtained with frequentist and Bayesian methods.

Our framework is illustrated in four case studies covering several different crops and pests: (1) *Septoria* leaf blotch (SLB) caused by the fungus *Zymoseptoria tritici* (Fuckel) on wheat (Eyal et al., 1987), (2) downy mildew caused by the oomycete *Plasmopara viticola* (Berk. & Curt.) on grapevine (Jermini et al., 2010), (3) yellow sigatoka caused by the fungus *Mycosphaerella musicola* R. Leach ex. J.L. Mulder on banana (Fouré, 1994), and (4) weevils (*Cylas formicarius*, Coleoptera) on sweet potato (Denon and Mauléon, 2004).

3.3 Materials and methods

3.3.1 General framework

We aimed to fit glmm models to data collected in regional surveys for monitoring the levels of a given pest or disease over the growing season in the region concerned. In the type of survey considered here, a pest or disease species is observed in N different site-years. Depending on the pest/disease species, the observations recorded (Y) may correspond to binary data (yes/no), disease incidence (percentage of diseased plants or organs), or count data (number of insects counted in traps). Observations are collected at each site-year on different dates during the growing season. Several observations are thus available in each site-year. In some surveys, additional data concerning sowing date, crop cultivar and risk level, for example, may be collected to characterise the site-years surveyed.

Our framework includes the following steps:

Definition of a glmm model. The characteristics of this model should be specified, taking into account the nature of the observations (binary data, disease incidence, or count data) collected in the regional survey considered.

At time t : fitting of the model defined in step (i) to the data collected up to time t . As explained below, the fitting procedure may be based on a classic statistical method or a Bayesian method. Use of the fitted model to estimate pest/disease dynamics up to time t , and to predict future changes in the pest/disease epidemic up to time $t+\Delta t$, where Δt represents a time interval (e.g., week, month).

Calculation of confidence or credibility intervals for analysis of the uncertainty of the estimated and predicted dynamics.

Steps (ii) to (iv) are repeated whenever new data become available during the growing season. Estimated and predicted dynamics can be displayed graphically, as shown in section 3.

3.3.2 Model specification

The glmm defined in step (i) is expressed as:

$$g(\mu_{ij}) = \alpha_i + \beta_i t_{ij} + \sum_{k=1}^K \gamma_k z_i^{(k)} \quad (6)$$

where μ_{ij} is the expected value of the j^{th} observation Y_{ij} collected in the i^{th} site-year at time t_{ij} , α_i and β_i are two site-year-specific regression parameters (varying across site-years), $z_i^{(k)}$, $k=1, \dots, K$, are K variables characterising the i^{th} site (e.g., risk level defined from sowing date and cultivar resistance), and γ_k are the K regression parameters associated with $z_i^{(k)}$, $k=1, \dots, K$.

In Eq.(6), μ_{ij} is related to the model input variables via a function g called the *link function*. Different link functions should be used for the different types of observations (Y) available. When Y_{ij} is a binary variable or corresponds to a proportion of infected plants, the logit function is a popular choice (Agresti, 2002 Chap.4). In this case, the function g is expressed as $g(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$. When Y_{ij} corresponds to count data, the log function is often appropriate, $g(\mu_{ij}) = \log(\mu_{ij})$ (Agresti, 2002 Chap.4).

Full specification of the model requires definition of the probability distribution of the observation Y_{ij} conditional to its expected value μ_{ij} , i.e., $P(Y_{ij}|\mu_{ij})$. Again, the choice here depends on the nature of the observations. For binary observations, $P(Y_{ij}|\mu_{ij})$ can be defined as Bernoulli. The binomial probability distribution is commonly used to analyse observations corresponding to proportions of infected plants, and the Poisson distribution is generally used to analyse count data (Agresti, 2002 Chap.4).

Eq.(6) assumes a linear relationship between $g(\mu_{ij})$ and time. This is realistic for monotonous pest-disease dynamics displaying an increase over the growing season, but is not appropriate for cyclic pest/disease dynamics. In such cases, the continuous variable t_{ij} can be replaced with a harmonic function combining sine and cosine (Prado and West, 2010 Chap.3). Another option is to replace t_{ij} with a factorial variable with several levels corresponding to different time periods, such as different months.

In glmm models, some parameters are considered to be random variables. In Eq.(6), α_i and β_i are indexed by the site-year index i . They are therefore assumed to vary across site-years according to a particular probability distribution. This distribution describes the between site-year variability of pest/disease dynamics. The use of random variables introduces greater flexibility than can be achieved with fixed parameter models. Thus, the model defined by Eq.(6) does not assume that all site-years follow the same dynamics. Instead, according to this model, pest/disease levels change more rapidly over the growing seasons in some site-years than in others, depending on the value taken by β_i .

The probability distribution of α_i and β_i can be defined in several different ways. One simple option is to define this distribution as a single joint Gaussian distribution $\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Sigma \right]$, where μ_α and μ_β are the expected values of α_i and β_i (i.e., their mean values over all site-years), and Σ is a variance-covariance matrix describing the variability of α_i and β_i around their mean values. μ_α , μ_β , Σ , and γ_k , $k=1, \dots, K$, are fixed parameters, the values of which must be estimated from the available data. More complex distributions can be defined. For example, it is possible to break α_i (or β_i) down into a sum of two terms corresponding to a random site effect and a random year effect, respectively, or into the sum of a random year effect and a random site-year effect. In these cases, a separate variance-covariance matrix should be specified for each set of random effects.

3.3.3 Case studies

3.3.3.1 Case study 1: *Septoria* leaf blotch in wheat

Data were collected in a survey conducted in two major wheat-producing regions in France (Champagne-Ardenne and Centre) (**Table 3.1**). On average, 78 wheat fields per year were selected from 2009 to 2015. SLB incidence was noted at least once per week for each field. At each date, the presence/absence of SLB symptoms was observed on the last three leaves of 20 different plants chosen at random from an 18 by 20 m plot of each selected wheat field that had been left untreated (Simmoneau et al., 2011b). Only incidences observed on the leaf L3 (the third leaf starting from the top) are considered here, but the same approach can be applied to other leaves. The total number of site-years was 934, corresponding to 5389 incidence data. Each site-year was characterised by a level of risk (low, medium, high) defined on the basis of wheat sowing date and the resistance of the wheat cultivar to SLB. These risk levels were defined by local experts.

These data were analysed with five glmm and five generalised linear models (glm) including only fixed parameters (**Table 3.2**). All these models considered the relationship between SLB and daily time. Some of these models included input variables indicating risk levels and/or regions. The glmm models included random effects describing between site-year variability. The glm models included no random effect.

3.3.3.2 Case study 2: Downy mildew on grapevine

Data were collected in the Midi-Pyrénées region of France from 2010 to 2015 (**Table 3.1**). The proportion of vines displaying downy mildew was determined annually, in a sample of vineyards, between April and August. Each vineyard was observed several times in one or several years. At least 50 vines were selected at random, from a row left untreated in each vineyard (IFV, 2010). The proportion of infected vines was determined by studying these selected vines. During the 2010-2015 period, 144 vineyards were monitored and 1782 proportions of infected vines were determined. On average, 25 vineyards were monitored and 130 data values were collected each year.

These data were analysed with five glmm and one glm (**Table 3.2**). All these models considered the relationship between mildew incidence and daily time. The five glmm different in terms of

their random effects. Three models included a single set of random effects corresponding to site effects, year effects, or site-year effects. Two glmm included two sets of random effects: site effects and year effects, or year effects and site-year effects. In this case study, it was possible to fit models including two sets of random effects because most of the vineyards were surveyed over several years.

3.3.3.3 Case study 3: Yellow sigatoka in banana

Data were collected from 282 plantations in Guadeloupe (French West Indies) over 11 years, from 2005 to 2015 (**Table 3.1**). The levels of yellow sigatoka infection on banana trees were assessed by calculating the stage of evolution of the disease (SED) index. This index was determined weekly, on 10 banana trees per site, according to the formula given by Ganry et al., (2008), taking into account the total number of leaves, the development stage of the disease in the five youngest leaves and the development stage of the cigar leaf of each tree.

We developed one glm and four glmm to determine the probability of the SED exceeding a critical threshold of 250 (**Table 3.2**). According to local experts, chemical treatments should be applied when this threshold is exceeded. Disease dynamics may follow seasonal cycles. The probability of $SED > 250$ was not, therefore, related to a continuous daily variable, but to a factor corresponding to monthly periods (the first five models) or to a harmonic function (the last model) (Prado and West, 2010 Chap.3). As for mildew (case study 2), glmm including a single set of random effect were compared with glmm including two sets of random effects (**Table 3.2**).

3.3.3.4 Case study 4: Weevils on sweet potato

Data were collected from 48 sites in Guadeloupe from 2011 to 2015. A trap containing pheromone was installed at each site one month after planting, and was checked for the presence of *Cylas formicarius* at fortnightly intervals until harvest (Champoiseau et al., 2011). The trapped insects were counted at each visit.

One glm and four glmm were developed to estimate the dynamics of weevil abundance. As in case study 3, time was expressed as a factor corresponding to monthly periods, and glmm including a single set of random effects were compared with glmm including two sets of random effects (**Table 3.2**).

Table 3.1: Main characteristics of the datasets

Disease/pest	Crop	Type of observation	Number of years	Number of site-years	Number of observations
<i>Septoria tritici</i> blotch	Wheat	Incidence (% of diseased leaves)	7	934	5389
Downy mildew	Grapevine	Incidence (% of diseased vines)	6	144	1782
Yellow Sigatoka	Banana	Binary (probability of exceeding a threshold)	11	282	6287
Weevils	Sweet potato	Count (number of insects)	5	48	277

3.3.4 Fitting procedures and model assessment

The parameters of the glm and glmm listed in **Table 3.2** were estimated by maximum likelihood, with the glm and glmer (lme4 package version 1.1-6) functions of R software (R Development Core Team 2014). We compared models by calculating the Akaike information criterion (AIC) (Akaike, 1974) and the root mean squared error (RMSE).

Median estimates were derived from $g^{-1} \left[\mu_{\alpha} + \mu_{\beta} t_{ij} + \sum_{k=1}^K \gamma_k z_i^{(k)} \right]$ where g^{-1} is the inverse link function. Site-year-specific estimates were derived from $g^{-1} \left[\alpha_i + \beta_i t_{ij} + \sum_{k=1}^K \gamma_k z_i^{(k)} \right]$. Median estimates are the disease/pest levels exceeded in 50% of the site-years. Site-year-specific estimates represent the pest or disease levels estimated for a given site-year. They were calculated for each site-year included in the datasets.

For glmm, two types of RMSE were calculated for each model: the RMSE values for median and site-year-specific estimates. For glm, RMSE was calculated for median estimates only, as no site-year-specific estimates were obtained for glm. The residuals obtained with the glm and glmm were analysed graphically.

The best model according to the AIC (i.e., the model with the lowest AIC) was selected and compared with its Bayesian version. Bayesian models were based on the same equations as those used in the frequentist models minimising AIC, but their parameters were estimated with a Markov chain Monte Carlo (MCMC) algorithm implemented with winBUGS software (Lunn et al., 2000). Weakly informative prior distributions were defined for the model parameters: $\mu_\alpha, \mu_\beta, \gamma_k \sim N(0, 10^{-4})$, $\Sigma \sim Wishart(\psi, \nu)$, with $\psi = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$, and $\nu = 2$.

The MCMC algorithm was run with 100,000 iterations. Convergence was checked by the Gelman-Rubin method (Gelman and Rubin, 1992). The first 50,000 iterations were discarded. Only one in 20 of the last 50,000 values was selected, to decrease the risk of autocorrelation.

The R and winBUGS codes are provided in **Annexe 7**.

3.4 Results

3.4.1 Septoria leaf blotch in wheat

Model 10septo had the lowest AIC and RMSE. This glmm includes risk level and region as covariables (**Table 3.2**). The AIC values obtained for the other models including random effects were all similar and at least 50% lower than the AIC values obtained for glm. This result is due to the strong variability of disease dynamics between site-years. Most of the RMSE for site-year-specific estimates derived from glmm were lower than the RMSE obtained for glm (**Table 3.2**). For example, the RMSE for model 10septo for site-year-specific estimates was 0.35 and was, therefore, much lower than the RMSE obtained for the glm 6septo (0.88). The residuals of the models were analysed graphically. The residuals were more symmetrically distributed around zero for glmm than for glm (not shown). Thus glm did not fit the data well due to strong between-site-year variability of disease dynamics.

Figure 3.1A shows the median outputs of the glmm 10septo (i.e., the disease levels likely to be exceeded in 50% of cases) for the Champagne-Ardenne and Centre regions and for the three risk levels. In each region, the median incidences estimated for sites characterised by high risk levels were higher than those estimated for sites with medium and low risk levels. Disease incidences were also higher in the Centre region than in Champagne-Ardenne (**Figure 3.1A**). These results

are consistent with the parameter values estimated for the different risk levels and the two regions (**Table 3.3**).

Figure 3.1B illustrates how a glmm can be used to estimate site-year-specific dynamics during the growing season. The fitted curve displayed in **Figure 3.1B** represents an example of fitted dynamic for a given site-year. This curve can be used to predict future disease incidences during a period of a few weeks after the last observation. As an illustration, in **Figure 3.1B**, the model is fitted to observations collected up to May 27th and used to predict disease incidence until June 20th. Disease dynamics can be updated whenever new observations are recorded for the site-year considered.

There was little difference between the 10septo model and its Bayesian variant. The parameter values estimated by the frequentist method were similar to the posterior medians obtained by the Bayesian method, and the confidence intervals were also similar to the Bayesian credibility intervals (**Table 3.3**). The only notable difference was that the estimated value of the Champagne-Ardenne parameter was slightly lower with the Bayesian method than with the frequentist approach. Consequently, for this region, estimated disease incidences were slightly lower with the Bayesian method (**Figure 3.2**). By contrast, the disease incidences estimated with the Bayesian and frequentist methods in the Centre region were indistinguishable (**Figure 3.2**). Quality of fit was similar for the two statistical methods, with an RMSE of 0.37 for the Bayesian method and 0.35 for the frequentist method. The Bayesian method has the advantage of being suitable for the calculation of credibility intervals for site-specific estimates, as illustrated in **Figure 3.1B**. Such calculations are not straightforward with frequentist methods. Bayesian methods thus provided a more complete picture of the uncertainty associated with estimates of disease dynamics.

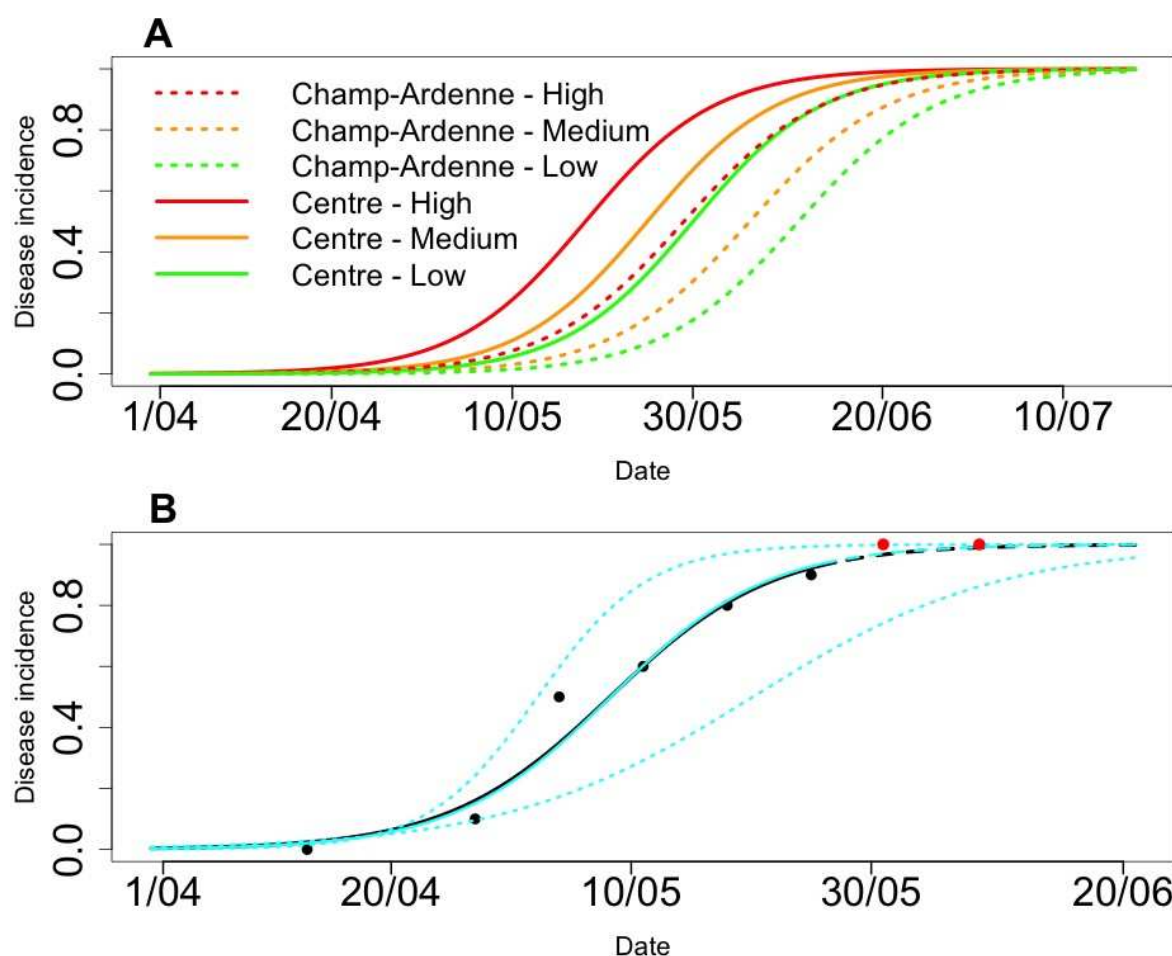


Figure 3.1. A. Median incidence of *Septoria* leaf blotch estimated for two French regions in 2015 — Champagne-Ardenne (dashed line) and Centre (solid line) — with the glmm 10septo. Three risk levels were distinguished: high (red), medium (orange) and low (green). B. Disease incidence data (points) and site-year-specific estimates obtained with the 10septo model for a given site-year in the Centre region in 2015. Site-year-specific estimates obtained before May 27th and after May 27th are indicated by solid and dashed lines, respectively. Observations collected before May 27th (black symbols) were used to fit the model and observations collected after May 27th (red symbols) were not used to fit the model. Frequentist estimates are indicated in black and Bayesian estimates in blue. Dotted blue lines represent Bayesian 2.5 and 97.5 percentiles.

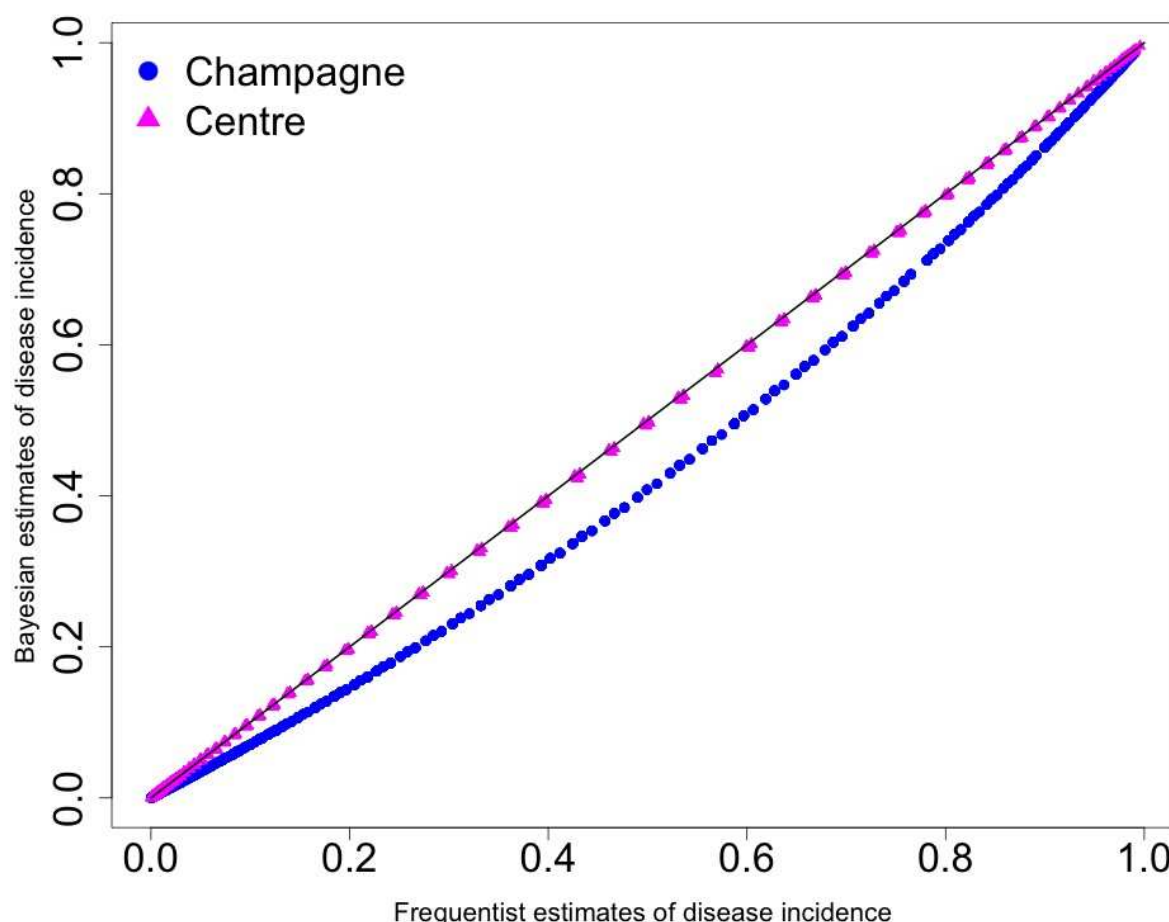


Figure 3.2: Bayesian estimates versus frequentist estimates of *Septoria* leaf blotch incidences calculated with the glmm model 10septo, for two regions (Champagne and Centre, represented by blue circles and magenta triangles, respectively).

3.4.2 Downy mildew in grapevine

The AIC of the glm was much higher than those of the glmm. The lowest values were obtained with the models including a single set of random site-year effects (5vine) or two sets of random effects (year effects and site-year effects; 6vine). Models 5vine and 6vine had the lowest RMSE values (0.46), and their residuals were more symmetrically distributed than those of the other models (**Figure 3.3**). These two models were thus considered the best for modelling downy mildew.

As the 6vine model includes year effects, it can be used to estimate disease dynamics year-by-year, as illustrated in **Figure 3.4A**. This figure shows the probability of vine infection as a

function of time for different years (from 2010 to 2015). Proportions of infected vines were estimated to be high in 2012 and 2013, and low in 2011 (**Figure 3.4A**). The 6vine model can also be used to estimate disease dynamics for specific site-years, as shown in **Figure 3.4B** for three different site-years corresponding to the same vineyard but to three different years (2011, 2013, and 2014). For the vineyard considered, proportions of infected vines were estimated to be low in 2011 and much higher in 2013 (**Figure 3.4B**). It was not possible to fit a Bayesian variant of this model due to its relatively high complexity. Convergence of the MCMC algorithm was not achieved for this model. However, we were able to fit the Bayesian variant of a model including random intercepts but a fixed time effect. The estimated parameters of this model differed slightly from those obtained for the frequentist version of the 6vine model (**Table 3.3**).

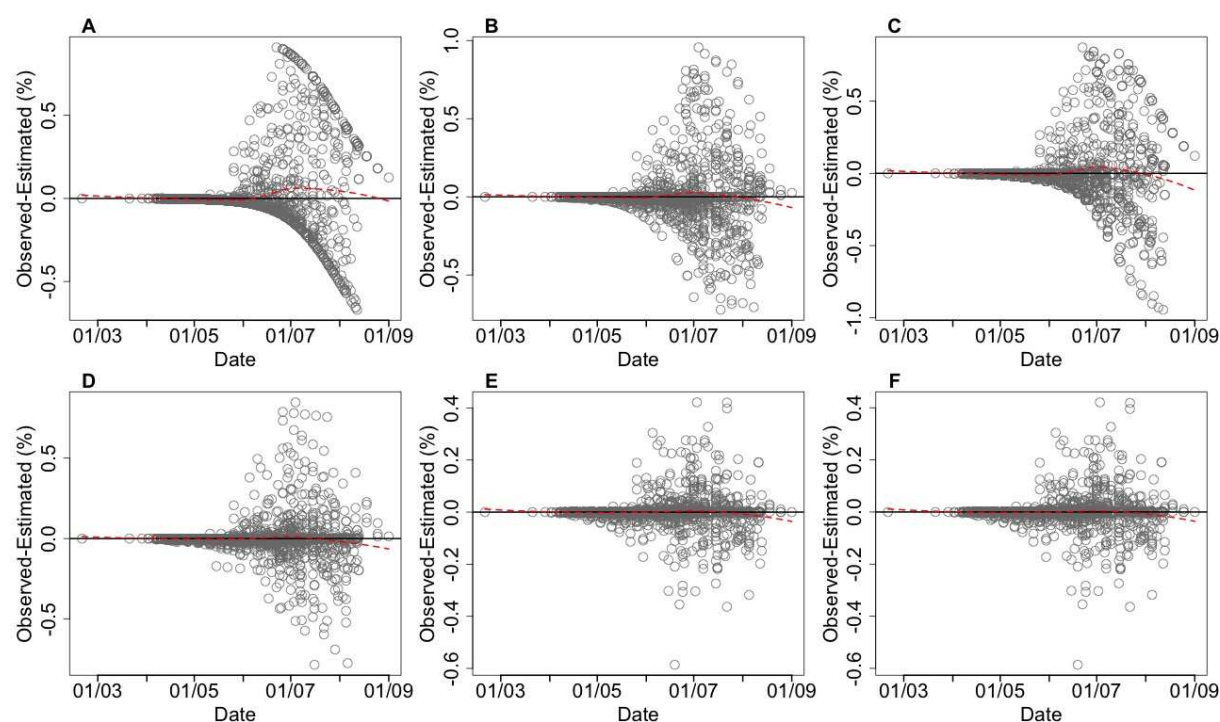


Figure 3.3: Residuals of the six models used to analyse downy mildew incidence on grapevine as a function of time. The dashed red curves were obtained by local polynomial regression (loess). A. Model 1vine. B. Model 2vine. C. Model 3vine. D. Model 4vine. E. Model 5vine. F. Model 6vine.

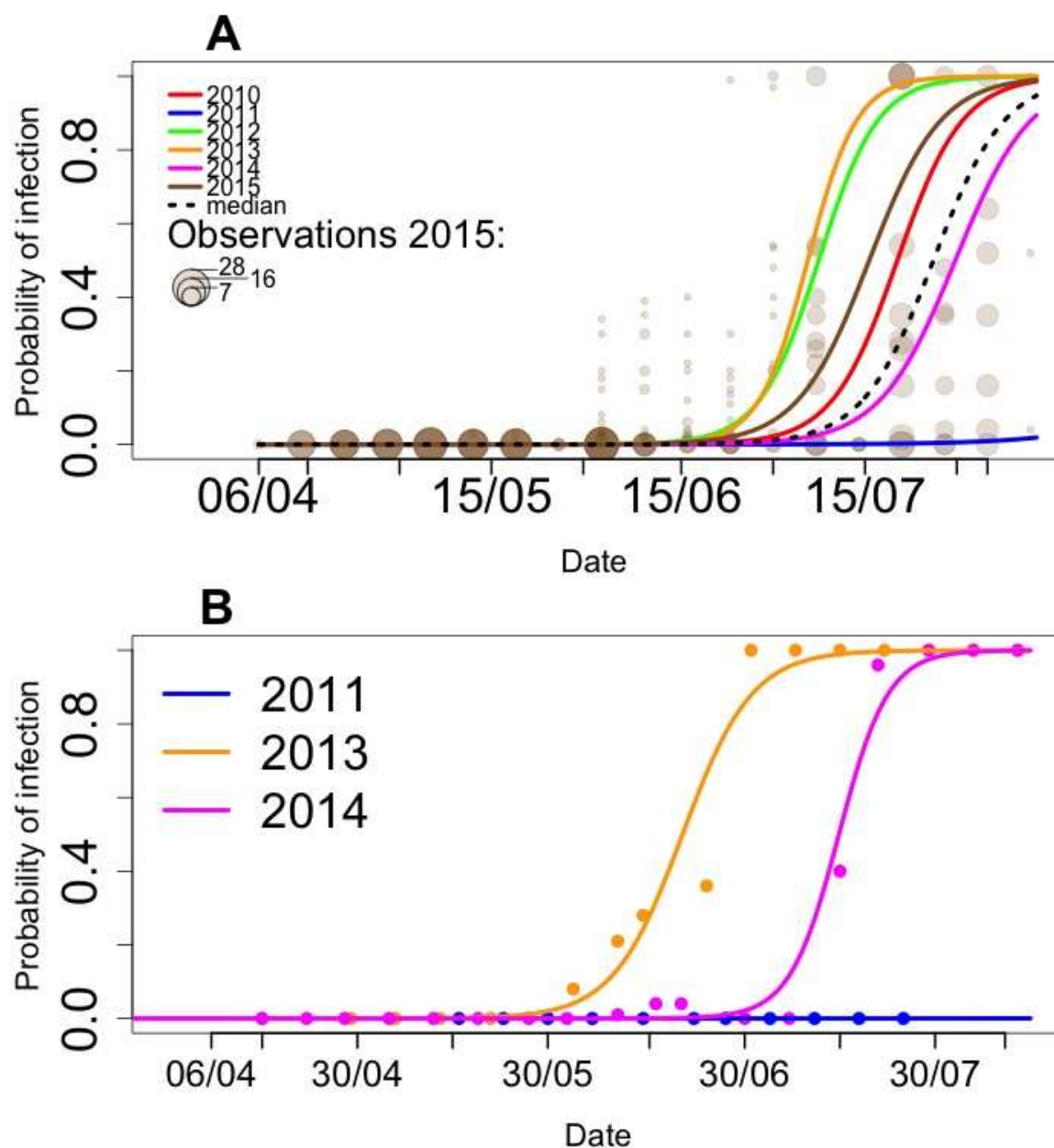


Figure 3.4 : A. Median disease incidence for downy mildew estimated from 2010 to 2015 with the 6vine model (for each year separately and over all years pooled together). Light brown circles indicate the observed data collected in the Midi-Pyrénées region in 2015. Circle size is proportional to the number of data. B. Disease incidence data (points) and estimated site-specific dynamics obtained with the 6vine model for a specific vineyard in 2011, 2013 and 2014.

Table 3.2: Characteristics of the models tested and values of several evaluation criteria: Akaike information criterion (AIC) and relative root mean square error (RMSE)

Model	Type of observation	Probability distribution	Link function	Time step	Random effect	Site characteristic	AIC	RMSE
1septo	incidence	Binomial	logit	days	none	none	63620	0.92
2septo	incidence	Binomial	logit	days	none	region	61486	0.90
3septo	incidence	Binomial	logit	days	none	risk level	61972	0.91
4septo	incidence	Binomial	logit	days	site-year	none	25513	0.99 (0.35)
5septo	incidence	Binomial	logit	days	none	region & risk level & region*risk level	59208	1.04
6septo	incidence	Binomial	logit	days	none	region & risk level	59324	0.88
7septo	incidence	Binomial	logit	days	site-year	region	25451	0.97 (1.02)
8septo	incidence	Binomial	logit	days	site-year	risk level	25467	0.97 (1.22)
9septo	incidence	Binomial	logit	days	site-year	region & risk level & region*risk level	25387	1.13 (0.35)
10septo	incidence	Binomial	logit	days	site-year	region & risk level	25384	0.94 (0.35)
1vine	incidence	Binomial	logit	days	none	none	55878	1.83
2vine	incidence	Binomial	logit	days	site	none	31742	1.95 (1.36)
3vine	incidence	Binomial	logit	days	year	none	41731	2.08 (1.61)
4vine	incidence	Binomial	logit	days	site & year	none	16769	1.98 (0.97)
5vine	incidence	Binomial	logit	days	site-year	none	6561	2.05 (0.46)
6vine	incidence	Binomial	logit	days	site-year & year	none	6516	2.05 (0.46)
1Sigatoka	binary	Bernouilli	log	month	none	none	6833	2.03
2Sigatoka	binary	Bernouilli	log	month	site	none	6352	2.05 (1.94)
3Sigatoka	binary	Bernouilli	log	month	year	none	6707	2.03 (2)
4Sigatoka	binary	Bernouilli	log	month	site & year	none	6232	2.05 (1.91)
5Sigatoka	binary	Bernouilli	log	month	site-year	none	6048	2.05 (1.8)
6Sigatoka	binary	Bernouilli	log	harmonic	site-year	none	6059	2.07 (2.09)
1weevils	count	Poisson	log	month	none	none	270584	2.28
2weevils	count	Poisson	log	month	site	none	108586	2.65 (1.65)
3weevils	count	Poisson	log	month	year	none	225364	2.30 (2.01)
4weevils	count	Poisson	log	month	site & year	none	92949	2.68 (1.42)
5weevils	count	Poisson	log	month	site-year	none	75534	2.68 (1.12)

3.4.3 Yellow sigatoka in banana

All glm had higher AIC values than glmm models. The inclusion of a seasonal effect (through a monthly time variable or a harmonic function) and of random site, year, or site-year effects decreased AIC values (**Table 3.2**). The 5Sigatoka model had the lowest AIC. This model includes a monthly time variable and random site-year effects. It also had a lower RMSE than the other models when this statistic was calculated from site-year-specific estimates (1.8) (**Table 3.2**). The residuals of the models were analysed graphically and appeared to be symmetrically distributed. The periods with the highest risk of infection were May, June, September, October, and November (**Figure 3.5A**). These peaks coincide with two different rainy seasons favouring spore dispersal (Rocha et al., 2012). Risks were lower in December-March and in July-August. This result was confirmed by the 6Sigatoka model, which included a harmonic function describing seasonality (**Figure 3.5A**).

Like the other glmm presented in the other case studies, the 5Sigatoka model can be used to obtain site-year-specific estimates of disease dynamics. **Figure 3.5B** shows the probability of exceeding an SED threshold of 250 for a given site-year. The highest probabilities were obtained in May-June and in September-November. This result is consistent with the observations; more than half the observed values of SED exceeded 250 during these two periods. The estimated parameters and the RMSE of the Bayesian variant of the 5Sigatoka model were very similar to those for the frequentist model (**Table 3.3**).

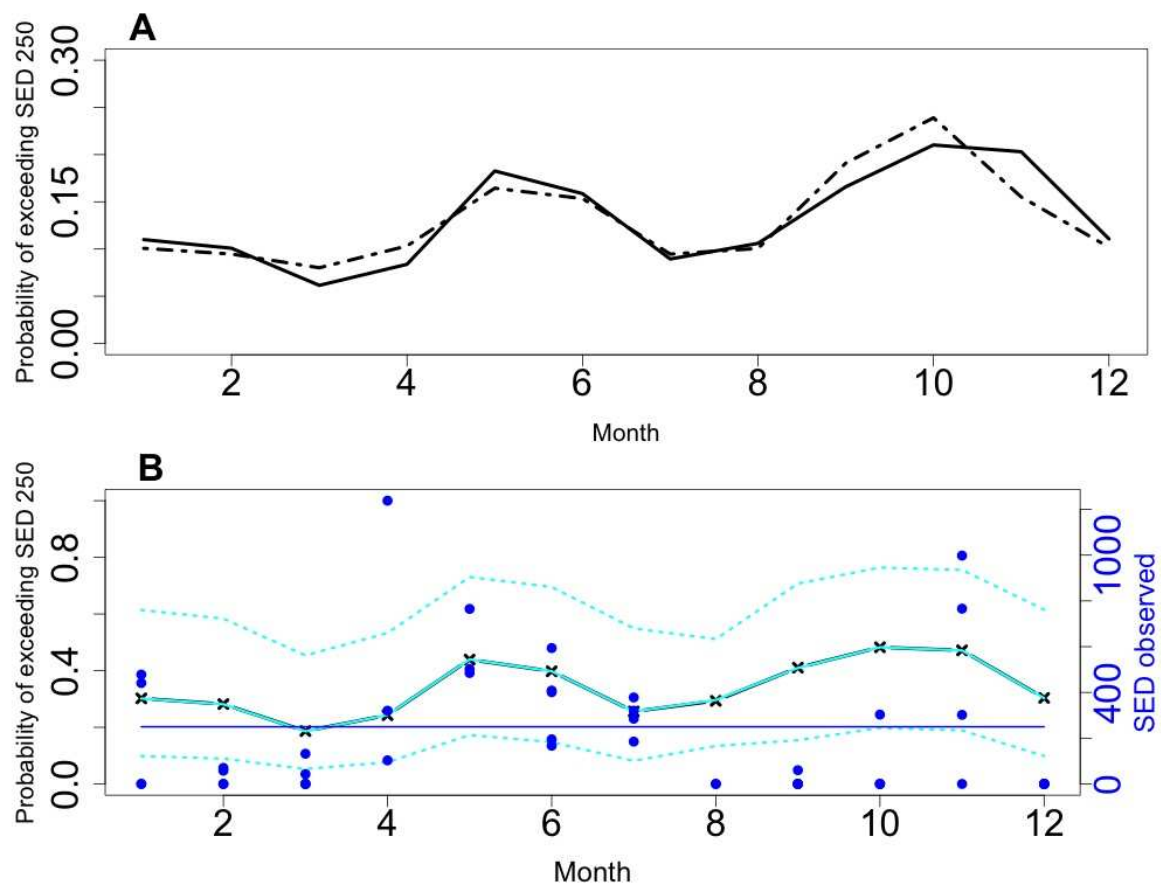


Figure 3.5: Probability of exceeding a threshold of 250 for the yellow sigatoka SED index, by month, in Guadeloupe. A. Median estimates of probability obtained with the 5Sigatoka model (continuous line) and with the 6Sigatoka model (dashed line). B. Estimated values of probability for a given site-year obtained with the 5Sigatoka model. Frequentist and Bayesian estimates are represented by black crosses and a continuous line, respectively (left scale). Dotted lines indicate the 2.5 and 97.5 percentiles of the estimated probabilities. The points are observed SED values (right scale) and the horizontal continuous blue line indicates SED=250.

3.4.4 Weevils on sweet potato

All the glmm had lower AIC values than the tested glm. The glmm with the lowest AIC was 5weevils (**Table 3.2**). This model included a monthly time variable and a site-year random effect. The lowest RMSE value (based on site-year-specific estimates) was also obtained for this model. The use of a

random year effect instead of a random site-year effect increased both AIC and RMSE substantially (**Table 3.2**).

The estimated median dynamics of weevil abundance obtained with the 5weevils model included two peaks, one in September and one in November (**Figure 3.6A**). This pattern may be accounted for by large numbers of weevils escaping waterlogged soils during the rainy season. These two peaks were even more marked for the third quartile of weevil abundance (**Figure 3.6A**). Weevil abundance was lowest in the December-June period.

An example of site-year-specific dynamics estimated by the 5weevils model and its Bayesian variant is shown in **Figure 3.6B**. The estimated abundance values obtained with the frequentist and Bayesian models were almost identical, and the estimated parameter values were similar (**Table 3.3**). Peaks of abundance were also visible in September and November in the site-year considered. The associated 95% credibility interval was very large, revealing a high level of uncertainty in the site-year specific estimates.

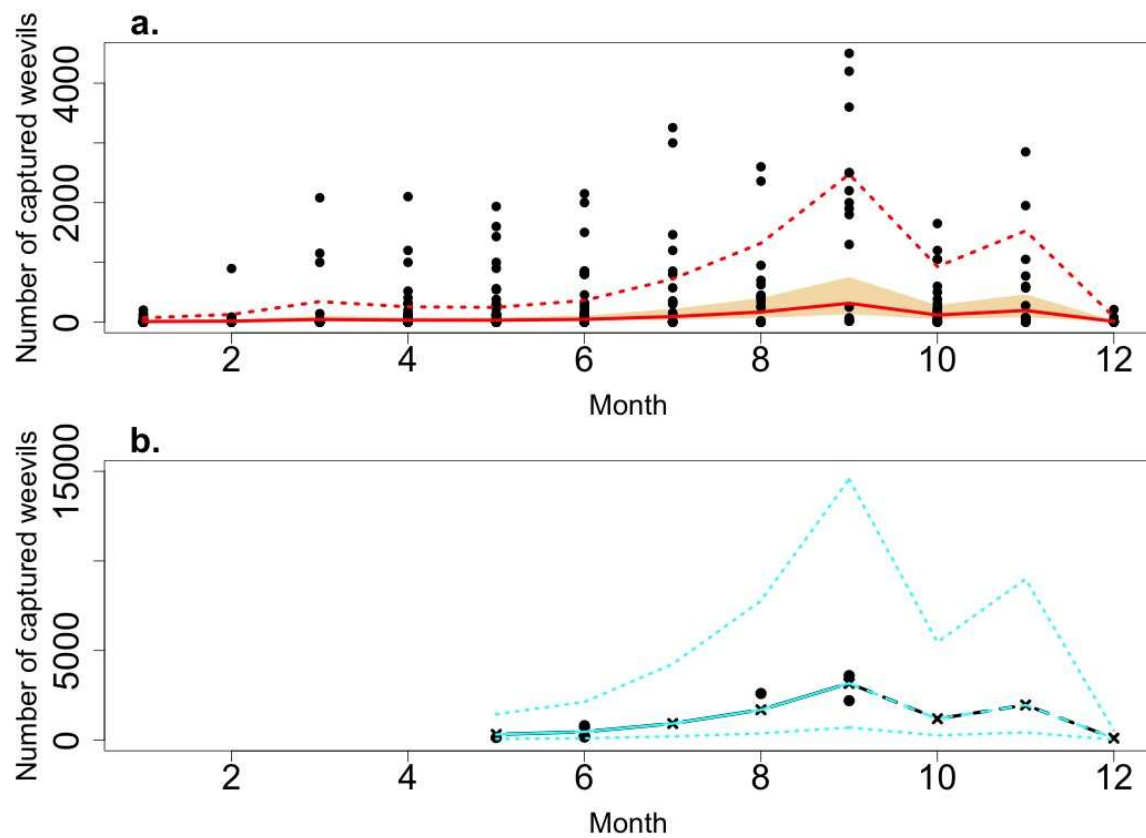


Figure 3.6: Estimated dynamics of weevil abundance in sweet potato crops in Guadeloupe, according to the 5weevils model. Black points indicate the observations. A. Median estimates of weevil abundance (red continuous line) and the 95% confidence interval (orange zone). Third quartiles of abundance are represented by the red dotted line. B. Dynamics estimated for a specific site-year with the frequentist version of the model (black crosses) and its Bayesian variant (blue lines). The dotted lines indicate the 95% Bayesian credibility intervals. Dashed lines indicate model predictions.

Table 3.3: Estimated parameter values (with confidence interval and credibility intervals) obtained with Frequentist and Bayesian methods for the selected models.

Model	Parameters	Frequentist	Bayesian
10septo	Centre*Low risk	-8.54 (-9.012 ; -7.9)	-8.538 (-8.737;-7.955)
	Time trend	0.14 (0.128 ; 0.15)	0.14(0.136;0.15)
	Champagne	-1.02 (-1.921 ; -0.54)	-1.427 (-1.523;-1.153)
	Time trend*Champagne	-0.008 (-0.019 ; 0.01)	-0.008 (-0.012;0.007)
	High risk	1.669 (1.229 ; 1.994)	1.682 (1.544;2.065)
	Medium risk	0.71 (0.332 ; 1.012)	0.716 (0.6;1.048)
6vine	Intercept	-9.23 (-11.075 ; -7.375)	-7.535 (-11.11 ; -6)
	Time trend	0.187 (0.115 ; 0.2)	0.134 (0.134 ; 0.136)
6Sigatoka	August	-2.132 (-2.441 ; -1.824)	-2.124 (-2.449 ; -1.811)
	April	-0.259 (-0.609 ; 0.091)	-0.27 (-0.619 ; 0.086)
	December	0.049 (-0.320 ; 0.419)	0.031 (-0.347 ; 0.424)
	Februrary	-0.056 (-0.428 ;0.317)	-0.073 (-0.455 ; 0.291)
	January	0.041 (-0.327 ; 0.41)	0.025 (-0.348 ; 0.418)
	July	-0.188 (-0.533 ; 0.156)	-0.203 (-0.554 ; 0.151)
	June	0.464 (0.136 ; 0.792)	0.451 (0.115 ; 0.78)
	May	0.634 (0.319 ; 0.949)	0.622 (0.299 ; 0.951)
	March	-0.593 (-0.974 ; -0.212)	-0.613 (-1.023 ; -0.231)
	November	0.765 (0.431 ; 1.1)	0.75 (0.41 ; 1.082)
	October	0.809 (0.477 ; 1.140)	0.795 (0.467 ; 1.129)
	September	0.517 (0.188 ; 0.847)	0.5 (0.166 ; 0.836)
5weevils	January	2.196 (3.071 ;1.32)	1.926 (1.399 ; 2.842)
	February	0.584 (0.679 ;0.49)	0.579 (0.485 ;0.683)
	March	1.583 (1.668 ;1.498)	1.577 (1.494 ;1.671)
	April	1.285 (1.368 ;1.201)	1.28 (1.195 ;1.367)
	May	1.252 (1.335 ;1.169)	1.248 (1.163 ;1.336)
	June	1.634 (1.718 ;1.551)	1.63 (1.545 ;1.72)
	July	2.326 (2.409 ;2.243)	2.321 (2.238 ;2.411)
	August	2.929 (3.012 ;2.845)	2.924 (2.84 ;3.015)
	September	3.558 (3.64 ;3.475)	3.553 (3.472 ;3.643)
	October	2.577 (2.661 ;2.492)	2.571 (2.488 ;2.663)
	November	3.075 (3.159 ;2.991)	3.07 (2.985 ;3.161)
	December	0.198 (0.328 ;0.068)	0.195 (0.0664 ;0.33)

3.5 Discussion

Our framework can be used to estimate the dynamics of pest and disease indicators (disease incidence, pest abundance, probability of exceeding a critical disease threshold) from regional surveys such as those conducted in France (MAAF 2008; MAAF 2015a). Our approach is based on the use of generalised linear mixed models (Bolker et al., 2009; Makowski et al., 2014; Piepho 1999). We demonstrate here that models of this type are flexible enough for adaptation to diverse crop species, pests and diseases. They can handle different types of data, including the binary data, incidence, and count data collected in regional surveys. They can also deal with repeated observations at the same site during the growing season, through the use of random effects. One or several sets of random effects can be included in these models to take into account one or several grouping factors, such as site, year, or site-year. Generalised linear mixed models can also include one or several covariables accounting for some of the variability in observations. In our case studies, the models considered included covariables relating to time (daily, monthly, seasonal), risk level and region. Other types of covariables, such as climatic variables, could be included, depending on the context. In case study 1, several glmm including climate variables (temperature sums and precipitation) were tested (not shown), but they performed no better than the other models and were therefore not selected in the end.

The proposed statistical models can easily be used to estimate the dynamics of pests and diseases at the regional level from regional survey data. They can be used to obtain a quantitative overview of the disease observations recorded in regional surveys. We believe that the models presented here have a role to play in alert systems for the monitoring of plant pest and disease risks. The results they generate could help farmers' advisors to formulate regional recommendations for disease and pest treatments. The estimated dynamics can be updated whenever new data become available (i.e., every week or every month, depending on the species), and plotted in the weekly alert bulletin published by the regional extension service. The estimated disease incidence dynamics can be compared with past disease dynamics, as shown for downy mildew in **Figure 3.4A**. The plotted dynamics can also be used to identify periods during which the risk of infestation is high, as illustrated for yellow sigatoka and weevils in **Figures 3.5** and **3.6**.

Another advantage of glmm is that they can be used to obtain site-specific estimates of pest and disease dynamics. Site-specific estimates can be used as a basis for local diagnosis and for the formulation of local recommendations. However, our results indicate that it is not always possible to obtain accurate local estimates if too few site-specific data are available.

We found almost no difference between frequentist and Bayesian models. Estimated parameter values and model outputs were very similar in most cases. These results are consistent with those obtained by Mila and Ngugi (2011) and Makowski et al., (2014). They are also consistent with those of Philibert et al., (2014) who found that their frequentist and Bayesian models had similar levels of accuracy. Bayesian models present the advantage of being able to describe the uncertainty associated with site-year-specific estimates through a posterior probability distribution (**Figures 3.1B-3.5B-3.6B**). However, these models require longer computation times than frequentist models. For example, estimation of the parameters of the 10septo model took 35 seconds with the frequentist approach and about an hour with the Bayesian approach (15 seconds vs. 55 minutes for 5Sigatoka, and 1 second vs. 2 minutes for 5weevils) on a computer equipped with an Intel® Core™ i7. The codes of our models are presented in **Annexe 7**. They could easily be adapted for the analysis of different types of pest and disease surveys.

3.6 Acknowledgements

We thank C. Abadie, L. Aurela, B. Cichosz, E. Gourdain, M. Héry, C. Jacoby-Koaly, F. Piraux, D. Simmoneau, and J.Veslot for their support and comments on our results and to Alex Edelman & Associates for English proofreading.

This research was supported by SynOEM - *Mieux profiter de la synergie entre réseaux d'observation, expertise et modélisation pour l'élaboration du Bulletin de Santé du Végétal* (<http://www.modelia.org/moodle/course/view.php?id=55>), funded by the French National Agency for Water and Aquatic Environments (ONEMA) under the Ecophyto Plan supported by the French Ministry of Agriculture. This work was also funded by a PhD grant from the French Association for Technical Research (ANRT).

Chapitre 4.

Étude de l'impact d'une réduction du nombre de parcelles observées sur la qualité des prédictions : application à la septoriose du blé en Champagne-Ardenne

4.1 Introduction

Dans le cadre de la surveillance biologique végétale, de nombreuses observations visuelles et/ou analyses en laboratoire sont effectuées afin de déterminer la présence ou l'absence d'un pathogène et d'estimer la prévalence d'une maladie ou d'un ravageur dans une région. Cette surveillance demande ainsi des moyens financiers et humains importants. La surveillance s'appuie sur un échantillon limité de parcelles prises dans une population globale (ensemble des parcelles d'une culture donnée dans une région donnée) ; chaque observation et analyse a un coût qui déterminera la taille de l'échantillon à analyser, le nombre d'unités (plantes) à prélever et observer ainsi que le nombre d'analyses en laboratoire à effectuer.

Dans un contexte de réduction budgétaire, il peut s'avérer utile de trouver une stratégie pour réduire la taille de l'échantillon tout en maintenant une précision des estimations aussi élevée que possible.

La région Champagne-Ardenne a été une des premières régions à mettre en place un réseau d'épidémiosurveillance à la suite du plan Écophyto (MAAF 2008). Celle-ci a débuté dès le printemps 2009, avec la mise en place d'un réseau d'observation sur le blé, puis s'est étendue en 2010 aux autres cultures de la filière Grandes Cultures et viticulture. Ce dispositif est présidé par la CRA – Champagne-Ardenne, supervisé par la DRAAF et piloté par un animateur inter-filière ; l'ensemble des acteurs techniques agricoles de la région font partie du dispositif d'épidémiosurveillance (instituts techniques : Arvalis dans le cas du blé, chambres régionale et départementales d'agriculture, coopératives, négoce, FREDON-Champagne-Ardenne) (DRAAF-SRAL 2010). Ainsi chaque année un nombre important d'observations est réalisé en région sur les parcelles de blé du réseau d'épidémiosurveillance.

Ce chapitre présente et compare différentes méthodes d'échantillonnage permettant de réduire le nombre de parcelles observées au sein du réseau de surveillance en cours de saison tout en conservant une qualité de prédiction satisfaisante du modèle présenté dans le chapitre 2.

Une étude sur l'optimisation de réseaux d'épidémiosurveillance a été réalisée par Arvalis – Institut du végétal en 2011 (Simmonneau et al., 2011a). Cette étude propose des abaques pour calibrer le réseau et ainsi déterminer le nombre optimum de parcelles à observer par région et par

bioagresseur afin de pouvoir analyser le risque. C'est sur la base des résultats de cette étude que la taille du réseau d'observation pour le blé en Champagne-Ardenne a été fixée à environ 70 parcelles observées chaque année. Ce nombre de parcelles suivies est fixe et toutes ces parcelles sont suivies du début à la fin de campagne régulièrement. Si des impasses existent (parcelles non observées certaines semaines), elles ne sont pas raisonnées pour optimiser l'échantillonnage, mais uniquement la conséquence de contraintes pratiques.

La plupart des travaux traitant de méthodes d'échantillonnage s'intéressent à l'échantillonnage intra-parcellaire (Delp et al., 1986 ; Disthaporn et al., 1993 ; Hughes et Gottwald, 1998). Certains de ces travaux permettent de générer une cartographie du développement de la maladie et d'analyser sa dispersion (Luo et al., 2012), d'autres permettent de détecter la maladie (Hughes and Gottwald, 1998 ; Ojiambo and Schern 2010) ou bien de développer et valider un plan d'échantillonnage. Par exemple, Perez et al., (2015) propose un plan d'échantillonnage pour estimer la densité de population de charançons sur canne à sucre.

Dans ce chapitre, nous considérons uniquement l'échelle régionale et nous comparons différentes méthodes d'échantillonnage permettant de réduire le nombre de parcelles observées au cours de saison sans trop diminuer la précision de nos prédictions. L'idée est d'utiliser un échantillon de N parcelles en début de saison, puis d'essayer de réduire ce nombre en cours de saison afin de diminuer les coûts, sans impact substantiel sur la précision des estimations d'incidence. Nous ne cherchons pas à modifier le protocole d'observation mais uniquement à réduire le nombre de parcelles observées en cours de saison.

Plusieurs méthodes d'échantillonnage sont comparées afin d'identifier celle qui est la plus adaptée pour réduire l'échantillon en cours de saison.

4.2 Matériel et méthodes

4.2.1 Description du dispositif actuel

En Champagne-Ardenne, le réseau d'observation pour le blé, qui a été mis en place à partir de 2009, est composé de 75 parcelles en moyenne chaque année (**Tableau 4.1**). Le nombre exact de parcelles varie chaque année.

Tableau 4.1: Nombre d'observations et de parcelles de blé observées pour chaque année et chaque feuille en Champagne-Ardenne.

	Nombre total de sites	Nombre de sites observés pour chaque feuille			Nombre d'observations pour chaque feuille		
		F3	F2	F1	F3	F2	F1
2009	81	80	80	78	513	471	418
2010	94	91	89	88	579	509	429
2011	70	68	64	64	384	318	284
2012	65	62	59	58	392	323	292
2013	72	71	71	70	449	398	353
2014	69	62	57	56	352	269	210

Le dispositif d'observation des parcelles est présenté en détail dans le chapitre 2. Ces parcelles sont choisies en début de campagne agricole (octobre) par les observateurs de manière à couvrir l'ensemble du territoire (localisation, variété). Dans le cas de la septoriose, ces parcelles sont observées au mieux une fois par semaine, du mois de mars jusqu'au mois de juin selon un protocole précis (Simonneau et al., 2011). Les parcelles sont classées en 3 groupes de risque (fort, moyen et faible) à partir des dates de semis et des variétés (résistance). L'incidence (présence/absence) de la septoriose est mesurée sur les 3 dernières feuilles (F1, F2 et F3) de 20 plantes échantillonnées dans une zone témoin non traitée de la parcelle.

Ces mesures d'incidence sont utilisées pour ajuster un modèle statistique par feuille qui estime la dynamique de l'incidence de la maladie à la parcelle pour les semaines suivantes. Ce modèle, décrit en détail dans le chapitre 2, est de type linéaire mixte généralisé et comporte des effets fixes des

groupes de risque et des effets aléatoires du site-année (un site-année étant la combinaison d'une année et d'une parcelle). Il permet d'estimer l'incidence de la maladie en cours de saison et de prédire les incidences futures pour les semaines suivantes à l'échelle de la région et à celle du groupe de risque. Les résultats de ce modèle sont inclus dans un rapport envoyé aux acteurs agricoles de la région afin de les aider à élaborer le BSV chaque semaine. La **Figure 4.1** présente des exemples de prédictions du modèle pour trois sites-années. Les données collectées avant la date du 19 mai servent à ajuster le modèle et faire les prédictions. Les données collectées entre le 19 mai et le 6 juin servent à évaluer l'erreur de prédiction.

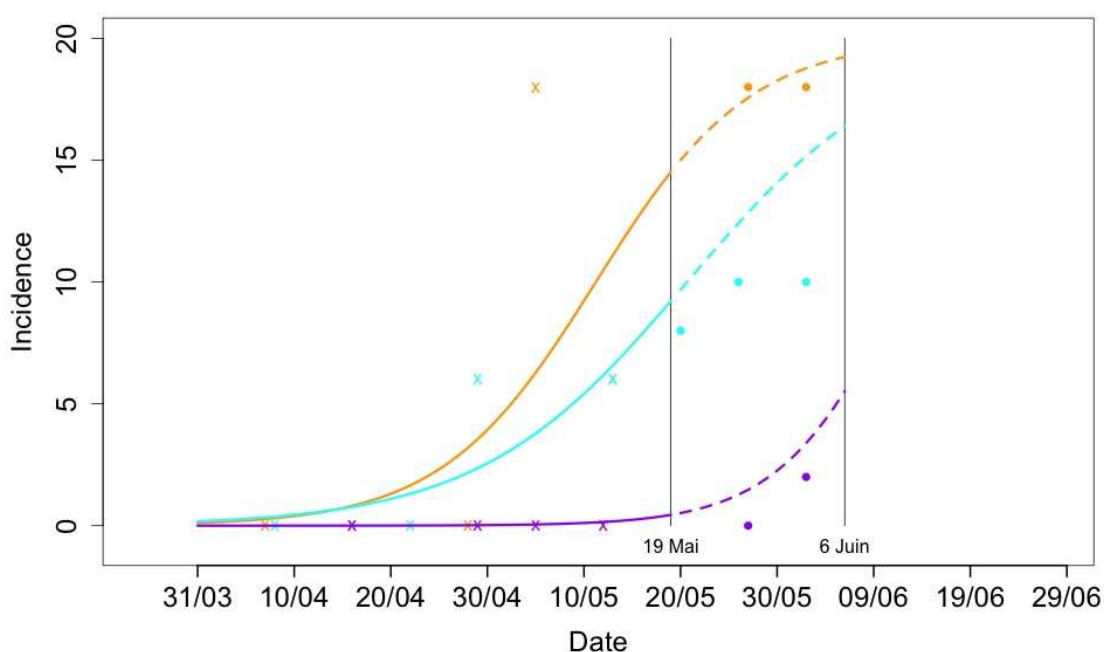


Figure 4.1 : Observations d'incidence de la maladie et estimations sites-spécifiques obtenue avec le modèle pour 3 sites-années différents en 2014 pour la F3. Les estimations sites-spécifiques obtenues avant le 19 mai et les prédictions obtenues après le 19 mai sont présentées en ligne continue et pointillée respectivement. Les observations collectées après le 19 mai (ronds) ne sont pas utilisées pour ajuster le modèle.

4.2.2 Description des stratégies d'échantillonnage testées pour réduire le nombre de parcelles suivies

L'échantillonnage se traduit ici par la sélection, à une date donnée en cours de saison, d'un sous-ensemble de parcelles par rapport à un ensemble initial de parcelles (Thompson 2012).

Nous avons défini deux dates d'échantillonnage SD_n (n allant de 1 à 2) du sous-ensemble de parcelles : le 1^{er} mai (SD1) et le 19 mai (SD2). L'incidence est estimée à 8 dates d'estimation dn (n allant de 1 à 8) : 7 mai (d1), 13 mai (d2), 19 mai (d3), 25 mai (d4), 31 mai (d5), 6 juin (d6), 12 juin (d7) et 18 juin (d8). Dans la suite du chapitre, lorsque nous échantillonnerons les parcelles à la date SD1 et SD2 nous estimerons l'incidence respectivement au 8 dates d'estimation pour SD1 et aux 5 dernières dates (de d4 à d8) pour SD2.

Nous testons 4 stratégies d'échantillonnage nommée : « 3 strates », « 7 strates », « groupe » et « référence » :

- « 3 strates » et « 7 strates » sont basées sur de l'échantillonnage par stratification, selon 3 et 7 strates respectivement, en définissant ces strates à l'aide des prédictions du modèle (voir ci-dessous pour les détails),
- « groupe » est basé sur de l'échantillonnage par stratification selon trois strates en définissant les strates à partir des trois groupes de risque des parcelles,
- « référence » est basé sur de l'échantillonnage aléatoire.

4.2.3 Mise en œuvre des stratégies « 3 strates » et « 7 strates »

L'échantillonnage par stratification comporte cinq étapes :

- Étape 1 : Ajustement du modèle et calcul d'incidence avant stratification
- Étape 2. : Stratification des parcelles à partir des estimations du modèle
- Étape 3 : Échantillonnage de parcelles dans chaque strate

- Étape 4 : Réajustement du modèle avec les observations collectées sur le sous-échantillon de parcelles
- Étape 5 : Calcul de l'incidence moyenne à partir du modèle et du sous-échantillon

Étape 1 : Ajustement du modèle et calcul d'incidence avant stratification

Les observations de l'année considérée antérieures à la date de stratification considéré SDn (1^{er} mai ou 19 mai) et l'ensemble des observations obtenues au cours des autres années sont utilisées pour estimer les paramètres du modèle statistique. Une fois les paramètres estimés, le modèle est utilisé pour estimer l'incidence de septoriose sur chacune des N parcelles de l'année en cours à la date SDn .

Étape 2 : Stratification des parcelles à partir des estimations du modèle

Les N parcelles sont ordonnées selon leur incidence estimée à la date SDn par ordre croissant, puis elles sont réparties en 3 strates (percentile 0.33 et 0.66 par rapport aux incidences estimées) ou en 7 strates (percentile 0.14, 0.29, 0.43, 0.57, 0.71 et 0.86 par rapport aux incidences estimées).

Étape 3: Échantillonnage de parcelles dans chaque strate

Un nombre identique de parcelles est choisi au hasard dans chaque strate de manière à constituer un sous-échantillon de taille $P < N$. Ce nombre est choisi de manière à réduire le nombre total de parcelles de l'échantillon de 25, 50, ou 75% par rapport au nombre initial N .

Étape 4: Réajustement du modèle avec les observations collectées sur le sous-échantillon de parcelles

Les paramètres du modèle sont ré-estimés à partir :

- des observations de l'année en cours collectées sur les N parcelles avant SDn ,
- des observations de l'année en cours des P parcelles sous-échantillonnées collectées avant la date dn ,
- et des observations collectées au cours des autres années.

Une fois les paramètres estimés avec l'ensemble de ces données, le modèle est utilisé pour estimer l'incidence de la septoriose sur les P parcelles du sous-échantillon.

Étape 5 : Calcul de l'incidence moyenne à partir du modèle et du sous-échantillon

Les valeurs d'incidence à la date dn sont estimées avec le modèle pour les P parcelles du sous-échantillon. Ces valeurs sont ensuite moyennées (mean_sampling).

La **Figure 4.2** illustre la technique de stratification et d'échantillonnage en prenant un exemple théorique de $N=9$ parcelles disponibles, « 3 strates » et une réduction d'1/3 des parcelles en cours de saison. Les 9 parcelles sont stratifiées en 3 strates avec 3 parcelles dans chaque strate à la date i , puis 2 parcelles sont choisies au hasard dans chaque strate. A la date $i+1$, des observations sont collectées sur les 6 parcelles du sous-échantillon.



Figure 4.2 : Exemple de stratification en 3 strates avec une sélection de 2 parcelles par strate

4.2.4 Comparaison des stratégies de stratification avec de l'échantillonnage aléatoire

La stratégie de stratification décrite ci-dessus est comparée à deux autres stratégies :

- Une autre stratégie de stratification (appelée « groupe ») dans laquelle trois strates sont définies à partir des trois groupes de risque des parcelles.
- Une stratégie « référence » basée sur de l'échantillonnage aléatoire sans stratification, où on tire aléatoirement dans l'ensemble des N parcelles pour constituer un échantillon de P parcelles.

Ces différentes stratégies de ré-échantillonnage sont comparées à l'aide d'une approche par bootstrap. Pour chaque stratégie, nous échantillonnons aléatoirement le nombre de parcelles P dans la population de parcelles disponibles et estimons l'incidence de ces P parcelles à la date dn , puis l'incidence moyenne (mean_sampling) est calculée (Étape 5). Cette approche est répétée 100 fois pour chaque année (2009 à 2014), chaque stratégie, chaque date dn et chaque type de feuille (F1, F2 et F3).

L'évaluation de chaque méthode d'échantillonnage (« 3 strates », « 7 strates », « groupe » et « référence ») s'effectue en calculant le Root mean square error (RMSE) à partir des différences entre les incidences moyennes obtenues par la stratégie d'échantillonnage (mean_sampling) et les incidences moyennes estimées avec l'ensemble des données collectées sur les N parcelles jusqu'à la date dn (mean_B).

Ce RMSE est calculé à chaque date, à la fois pour chaque année et pour toutes années confondues.

4.2.5 Exemple illustré de stratification et d'échantillonnage

Dans cette partie, nous illustrons la démarche en prenant l'exemple de la stratégie d'échantillonnage « 3 strates », pour la feuille 3 et l'année 2014 avec une date de stratification au 19 mai (SD2), un pourcentage de réduction de 75% et une estimation au 31 mai (d5).

Étape 1: Les observations de 2014 antérieures au 19 mai (croix vertes de la **Figure 4.3**) ainsi que l'ensemble des observations des autres années sont utilisées pour estimer les paramètres du modèle.

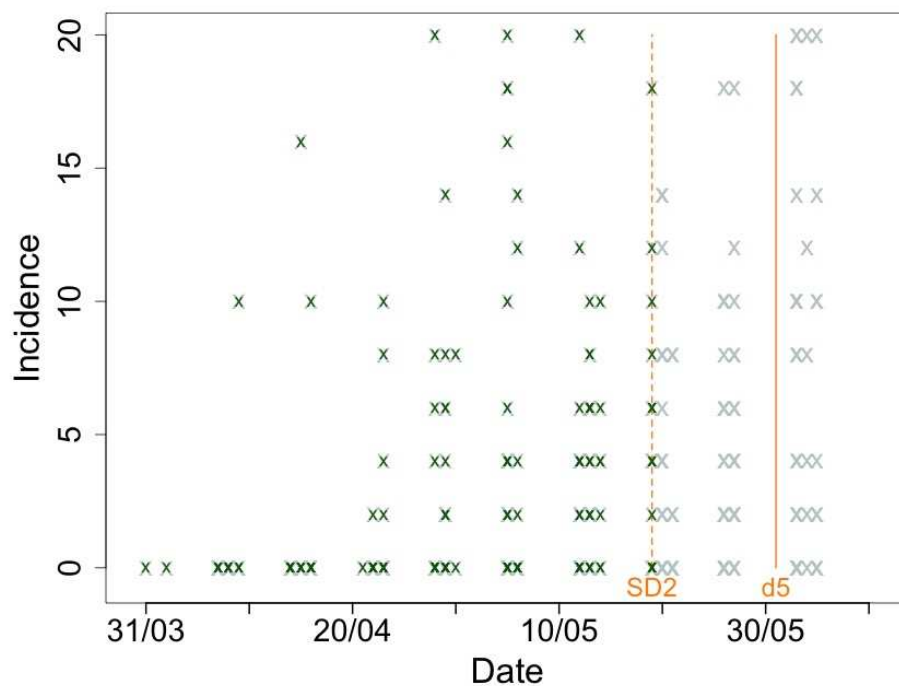


Figure 4.3 : Observations d'incidence en fonction de la date, l'ensemble des observations de 2014 sont représentées par les croix, les observations utilisées pour ajuster le modèle avant stratification sont en vert.

Une fois les paramètres estimés, nous calculons l'incidence de septoriose sur chacune des N parcelles ($N=62$) de 2014 à la date du 19 mai (courbes vertes pointillées de la **Figure 4.4**)

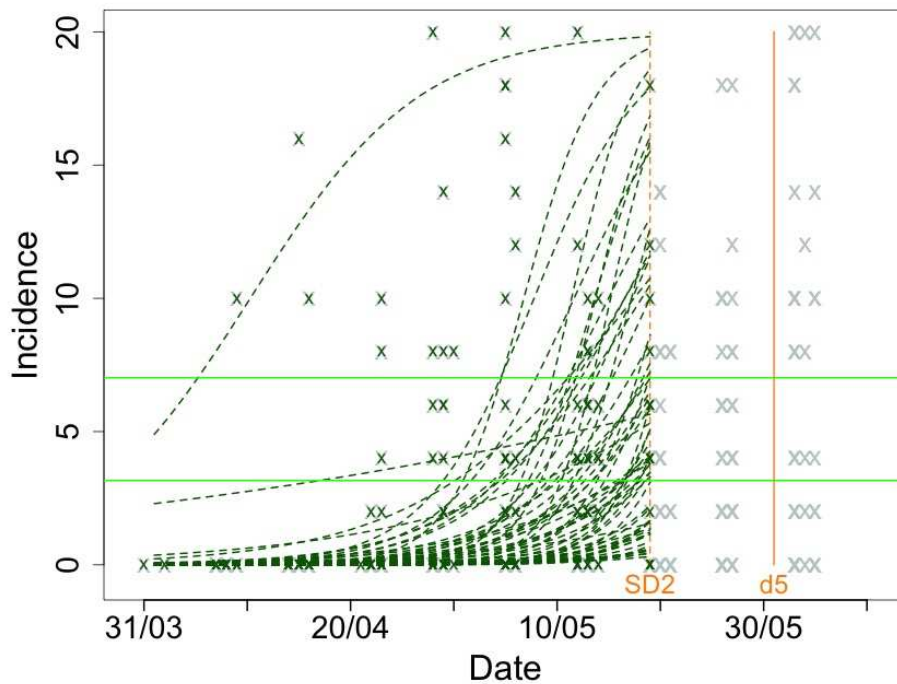


Figure 4.4 : Incidences au 19 mai de l'ensemble des parcelles de 2014 estimées avec le modèle, représentées en vert pointillé.

Étapes 2: Les 62 parcelles disponibles sont réparties en 3 strates de même taille (groupes définis par les traits horizontaux verts de la **Figure 4.4**).

Étapes 3: Dans notre exemple, au 19 mai il y a donc 62 parcelles disponibles pour l'échantillonnage ; pour une réduction de 75%, nous échantillonnons au hasard 5 parcelles dans chaque strate ($P=15$).

Étape 4: Les observations de 2014 collectées sur les 62 parcelles avant le 19 mai (croix vertes sur la **Figure 4.5**), les observations des 15 parcelles sous-échantillonnées collectées avant le 31 mai (carré vert sur la **Figure 4.5**) et l'ensemble des observations collectées au cours des autres années sont utilisées pour ré-estimer les paramètres du modèle.

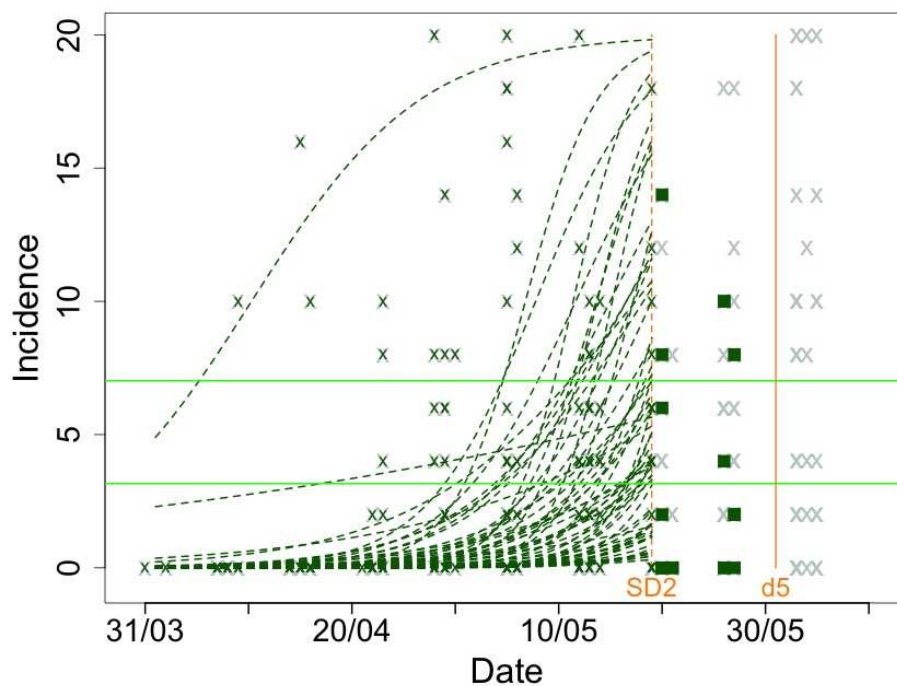


Figure 4.5 : Observations des 15 parcelles sous-échantillonnées entre le 19 et 31 mai 2014 sont représentées par les carrés verts.

Étape 5 : Nous estimons alors l'incidence de chacune des 15 parcelles du sous-échantillon à la date du 31 mai (courbe pleine verte de la **Figure 4.6**) ainsi que l'incidence moyenne de ces parcelles qui correspond à mean_sampling (rond vert).

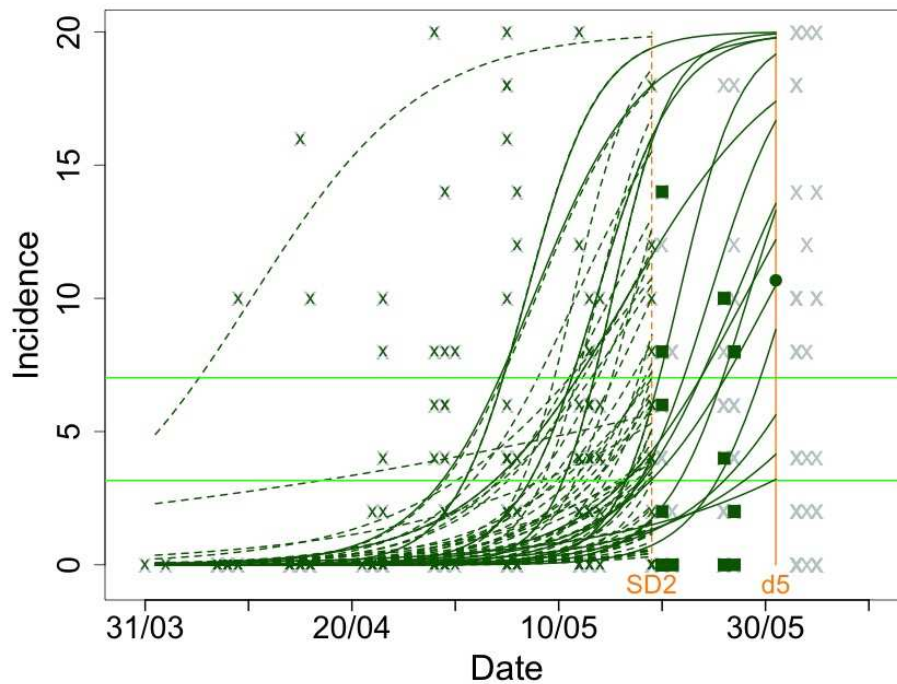


Figure 4.6 : Incidences au 31 mai des parcelles échantillonnées (courbes vertes continues) et l'incidence moyenne de ces parcelles (rond vert).

Pour l'évaluation de la stratégie de cet exemple, cette moyenne (mean_sampling) est à comparer à l'incidence moyenne estimée avec l'ensemble des données collectées sur les 62 parcelles jusqu'à la date du 31 mai (mean_B).

Les étapes 3, 4 et 5 représentées par les figures **Figure 4.5** et **Figure 4.6** sont répétées 100 fois, ainsi on aura 100 valeurs de mean_sampling ce qui permettra de calculer le RMSE en faisant la racine carrée de la somme des carrés, des différences entre les 100 mean_sampling (rond vert) et mean_B.

4.3 Résultats et discussion

La méthode qui conduit le plus souvent aux RMSE les plus faibles est la stratégie de stratification «7 strates», suivie de «3 strates» et de «groupe» (**Tableau 4.2**). Les performances de l'échantillonnage non stratifié («référence») sont toujours moins bonnes.

Tableau 4.2: RMSE (%) des 3 types de feuilles et 6 années pour chaque stratégie d'échantillonnage et % de réduction des parcelles. Les RMSE notées en rouge sont les plus faibles pour chaque % de réduction du nombre de parcelles. Les RMSE ont été moyennées sur toutes les dates SDn et dn. La colonne « Global » correspond aux RMSE moyennés sur les trois types de feuilles et les six années.

% de réduction	Stratégie	Global	F3	F2	F1	2009	2010	2011	2012	2013	2014
75	Référence	6.83	5.93	7.6	7.39	4.74	3.04	3.11	9.58	6.74	8.8
75	3 strates	5.98	5.16	4.34	8.55	4.56	3.14	2.55	6.16	5.12	9.25
75	7 strates	4.98	5.46	4.65	3.44	4.21	1.93	2.92	6.79	5.4	5.74
75	groupe	6.8	6.3	4.95	9.3	5.04	3.09	3.13	6.77	7.04	10.38
50	Référence	5.09	4.2	5.77	5.68	2.98	1.99	1.91	7.6	5.37	6.27
50	3 strates	4.16	3.59	2.95	5.99	2.9	1.97	1.85	4.48	3.66	6.44
50	7 strates	3.29	3.88	2.73	1.81	2.93	1.23	1.15	4.56	2.21	3.78
50	groupe	4.87	4.51	3.57	6.68	3.17	2.44	2.23	5.4	5.54	6.76
25	Référence	4.35	3.5	4.85	5.12	2.11	1.36	1.47	6.64	4.94	5.26
25	3 strates	3.69	2.78	2.33	5.71	1.89	1.42	1.13	3.92	3.45	5.76
25	7 strates	2.02	2.11	1.94	1.66	1.95	0.9	1.29	2.49	1.92	2.54
25	groupe	6.03	2.69	2.44	9.18	NA	NA	NA	1.96	2.15	7.06

Lorsque l'on regarde feuille par feuille, nous remarquons que pour la F1 la stratégie qui a le plus petit RMSE est «7 strates» (quelques soit le pourcentage de réduction de parcelle). Pour la F2, la stratégie «7 strates» a le plus faible RMSE lorsque l'on prend la réduction de 75% et 50% des parcelles, dans le cas de la réduction à 25% des parcelles c'est la stratégie «3 strates». Pour la F3, la stratégie «7 strates» a le plus petit RMSE lorsque l'on prend la réduction de 25 % des parcelles, dans le cas de la réduction à 50 et 75% des parcelles c'est la stratégie «3 strates». Les RMSE sont plus grands pour la F3 et pour la F2 et que pour la F1, ce qui est présenté également sur la **Figure 4.7**.

En ce qui concernent les années, la stratégie « 7 strates » et « 3 strates » ont les RMSE les plus faibles quelques soit le pourcentage de réduction de parcelles (hormis pour la réduction de 25% des parcelles en 2012 où c'est la stratégie «groupe » qui a le meilleur RMSE).

Pour les années 2010 et 2014, la meilleure stratégie est dans tous les cas « 7 strates ». Pour les autres années, les RMSE entre la stratégie « 3 strates » et « 7 strates » sont extrêmement proches.

Dans l'ensemble les RMSE sont assez proches entre les années, seules les années 2012 et 2014 ont des RMSE plus élevés que les autres années de l'ordre de 1 à 6%.

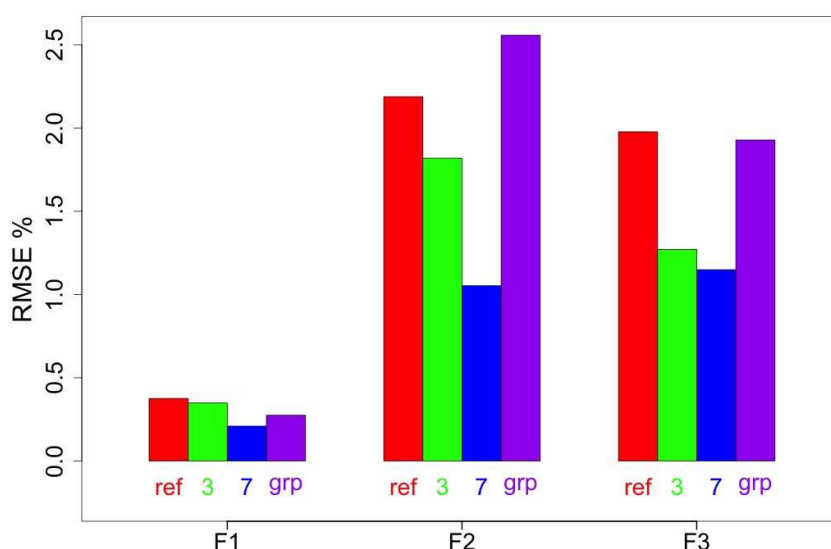


Figure 4.7 : RMSE (%) moyens sur toutes les années des 4 stratégies d'échantillonnage pour chacune des 3 feuilles (« référence » en rouge, « 3 strates » en vert, « 7 strates » en bleu et « groupe » en violet). Les RMSE sont calculés pour une date de stratification au 19 mai, une date d'estimation au 18 Juin et un pourcentage de réduction des parcelles de 50%.

Le **Tableau 4.3** montre que quelques soient la date de stratification (SD1 ou SD2), les RMSE les plus faibles sont obtenus pour la stratégie « 7 strates ».

Plus la date d'estimation est éloignée (de d1 à d8 pour SD1 et de d5 à d8 pour SD2), plus le RMSE (quelque soit la stratégie adoptée) est élevé. Ce constat est illustré avec la **Figure 4.8**.

L'augmentation du RMSE avec la date d'estimation quelques soient la date de stratification est due au fait que l'incidence de la maladie augmente.

Tableau 4.3: RMSE (%) des 2 dates de stratification et pour toutes les dates d'estimation selon la date de stratification pour chaque stratégie d'échantillonnage et % de réduction des parcelles. Les RMSE notés en rouge sont les plus faibles pour chaque % de réduction en colonne. Les RMSE ont été moyennés sur les types de feuille et les années.

Stratégie	% de réduction			SD=62 (SD1)								SD=80 (SD2)				
		SD=62 (SD1)	SD=80 (SD2)	d1	d2	d3	d4	d5	d6	d7	d8	d4	d5	d6	d7	d8
75	Référence	8.78	4.43	2.32	2.49	2.98	4	6.44	9.36	13.5	16.41	2.73	3.11	3.91	4.79	6.54
75	3 strates	7.78	4.27	2.03	2.37	2.89	4.07	6.61	10.2	11.73	12.79	2.05	2.43	3.49	4.7	6.84
75	7 strates	6.15	4.22	2.42	3.06	3.78	4.43	5.83	7.72	8.87	9.02	2.08	2.53	3.6	4.76	6.52
75	groupe	8.8	4.63	2.51	2.77	3.25	4.54	7.3	11.49	13.17	14.7	2.61	2.92	3.96	5.07	7.11
50	Référence	6.73	2.93	1.51	1.67	2.12	2.8	4.37	6.74	10.46	13.09	1.72	1.89	2.42	3.13	4.56
50	3 strates	5.63	2.68	1.44	1.73	2.25	3	4.81	7.32	8.47	9.25	1.3	1.46	2.14	2.9	4.35
50	7 strates	4.69	2.4	1.8	2.37	2.94	3.73	4.8	6.1	6.77	6.22	0.99	1.35	2.2	2.82	3.62
50	groupe	6.1	3.27	1.58	1.83	2.25	3.12	4.99	7.63	9.11	10.53	1.86	1.92	2.48	3.41	5.33
25	Référence	5.93	2.12	0.86	1.12	1.57	2.22	3.64	5.87	9.29	11.75	1.12	1.13	1.54	2.14	3.6
25	3 strates	5.36	2.05	0.9	1.18	1.75	2.46	4.29	6.8	8.28	9.22	1.05	0.95	1.37	2.04	3.59
25	7 strates	2.68	1.73	1.29	1.78	2.45	3.02	3.28	2.84	3.06	3.05	0.81	0.81	1.22	1.9	2.91
25	groupe	7.06	2.06	1.43	1.63	1.86	2.79	5.23	9.31	10.86	12.28	0.53	0.89	1.3	2.05	3.76

Pour SD1, pour les premières dates d'estimation, les stratégies d'échantillonnage ne se distinguent que très peu (avec un faible avantage pour la stratégie « référence » lorsque l'on réduit de 50 et 75% le nombre de parcelles et pour la stratégie « 3 strates » lorsque l'on réduit de 25%). Cependant, après d5 c'est la stratégie « 7 strates » qui a le plus petit RMSE.

Pour SD2, de d4 à d7, lorsque l'on réduit de 75% le nombre de parcelles, la stratégie « 3 strates » a le plus petit RMSE. Lorsque l'on réduit le nombre de parcelles de 25% c'est la stratégie « 7 strates » qui a le RMSE le plus faible. Pour une réduction de 50% des parcelles, les deux stratégies sont très proches.

La stratégie « groupe » n'a jamais le plus petit RMSE (sauf dans le cas d'une réduction 25% des parcelles en 2012).

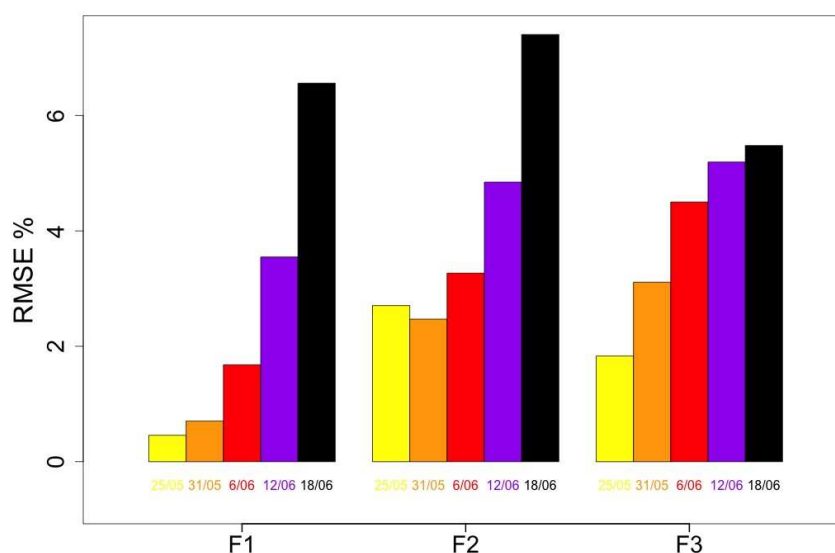


Figure 4.8 : RMSE (%) moyens sur toutes les années pour 5 dates d'estimation pour chacune des 3 feuilles (25 mai en jaune, 31 mai en orange, 6 juin en rouge, 12 juin en violet et 18 juin en noir). Les RMSE sont calculés pour une date de stratification au 19 mai, la stratégie d'échantillonnage « 7 strates » et un pourcentage de réduction des parcelles de 75%.

Le RMSE est plus faible lorsque le pourcentage de réduction est de 25% (**Tableau 4.2**, **Tableau 4.3** et **Figure 4.9**). Ce résultat est logique ; plus la taille de l'échantillon est réduite, plus l'erreur d'estimation augmente.

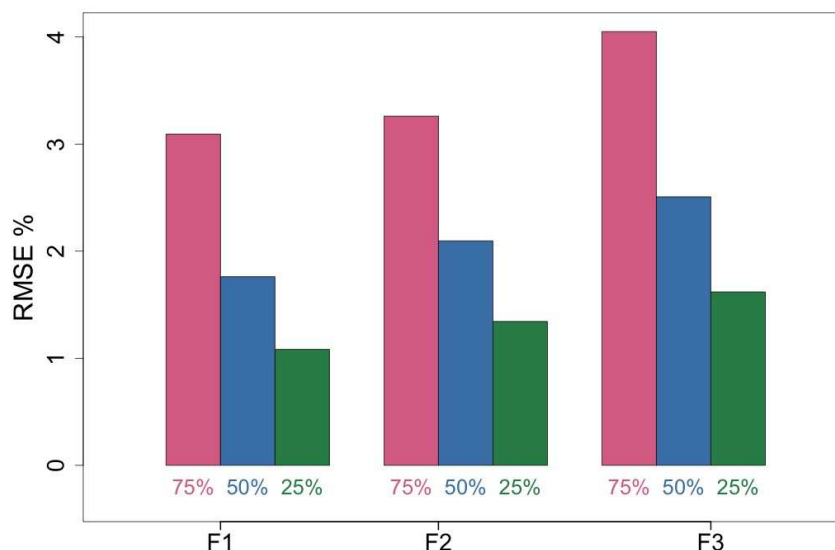


Figure 4.9 : RMSE (%) global (toutes années confondues) des 3 pourcentages de réduction de parcelles pour chacune des 3 feuilles (25% de réduction en vert, 50% de réduction en bleu et 75% de réduction en rose). Les RMSE sont calculés pour une date de stratification au 19 mai, une date d'estimation au 25 mai et la stratégie d'échantillonnage « 3 strates ».

L'ensemble des calculs de RMSE réalisés montre l'intérêt d'effectuer une stratification pour réduire la taille de l'échantillon. Les RMSE des stratégies « 3 strates » et « 7 strates » sont dans la plupart des cas les plus petits. Ces deux stratégies sont donc à privilégier lorsque l'on souhaite réduire la taille de l'échantillon de parcelles en cours de saison. Leurs performances s'expliquent par l'utilisation d'un modèle statistique permettant d'estimer l'incidence de la maladie pour les N parcelles de l'échantillon initial. Ces estimations permettent de définir des strates relativement homogènes et, ainsi, d'optimiser le sous-échantillonnage.

La **Figure 4.10** présente de manière synthétique pour la F3, les RMSE des 4 stratégies étudiées pour les 3 pourcentages de réduction du nombre de parcelles à la date de stratification du 19 mai (SD2) et la date d'estimation du 31 mai (d5). Selon le pourcentage d'erreur visé (RMSE), la stratification permet de réduire plus ou moins fortement le nombre de parcelles observées en cours de saison. Par exemple au 19 Mai (SD2), en réduisant de 50% le nombre de parcelles observées et

en stratifiant selon « 7 strates », l'erreur sera de 1.35% (soit 0.27 feuille sur 20) si l'on estime l'incidence au 31 Mai (d5).

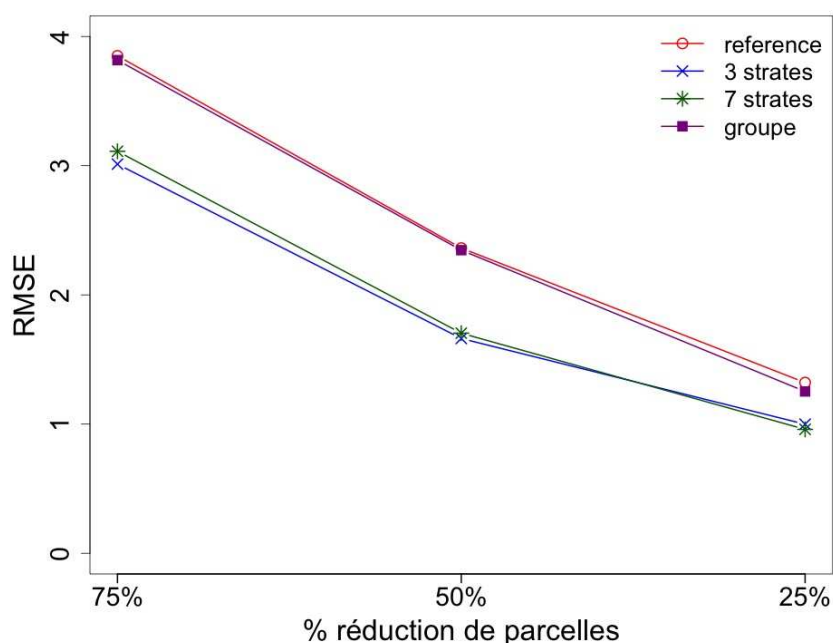


Figure 4.10 : RMSE (%) pour la F3 pour les 4 stratégies d'échantillonnage en fonction des 3 niveaux de réduction de parcelles RMSE calculées avec une date de stratification au 19 mai (SD2) et une date d'estimation au 31 mai (d5). RMSE moyennés sur les années.

L'intérêt de réduire le nombre de parcelles observées afin de diminuer les coûts d'observation a été discuté lors des divers échanges avec les acteurs du BSV. Dans le cas d'une réduction du nombre de parcelles en cours de saison, la méthode d'indemnisation des partenaires observateurs et les règles de suivi pourrait être revue.

Nous aurions pu stratifier directement à l'aide des observations du réseau d'épidémiosurveillance, cependant le faible nombre d'observations est une contrainte (une parcelle est observée en moyenne six fois lors de la saison), l'utilisation du modèle permet potentiellement de stratifier à un plus grand nombre de date étant donné que l'on estime la dynamique de l'incidence sur l'ensemble de la saison.

Il est important de noter que les parcelles de blé suivies dans le réseau ne sont pas observées uniquement pour la septoriose mais également pour d'autres maladies (fusariose, rouille brune/jaune, oïdium, piétin-verse), ravageurs et le suivi physiologique (stade, gel,...). La méthodologie d'échantillonnage proposée pourrait être adaptée à ces autres maladies afin de confirmer la possibilité de réduction du nombre de parcelles en cours de saison.

Nous avons choisi de sélectionner nos parcelles de manière aléatoire, nous n'avons pas regardé d'autre possibilité comme celle d'échantillonner de préférence les parcelles les plus accessibles demandant le moins de déplacement, critère pratique également essentiel à prendre en compte.

D'autres stratégies d'échantillonnage pourraient être analysées en jouant, en particulier, sur la fréquence des observations ou bien le nombre de témoins au sein d'une même parcelle ou encore le nombre de feuilles à échantillonner au sein de chaque témoin.

Chapitre 5.

Co-construction et mise en œuvre opérationnelle de l'outil

Ce dernier chapitre présente la mise en œuvre opérationnelle de la démarche et de l'outil statistique développé pour le cas d'étude de la septoriose du blé auprès des acteurs du BSV en région. Après une série de test en conditions réelles, l'outil est devenu opérationnel dès la campagne 2016 et est dorénavant utilisable directement par les acteurs des deux régions Champagne-Ardenne et Centre.

5.1 Une thèse « appliquée » : co-construction d'un outil avec les acteurs du BSV

Cette thèse s'inscrit dans le projet de recherche PSPE dénommé SynOEM, ce projet est multi-partenarial dans la mesure où l'ensemble des acteurs du BSV pour chaque cas d'étude sont parties prenantes du projet. Ainsi les partenaires ont collaboré à la construction de l'outil opérationnel résultant de ces travaux de thèse qui revêt ainsi le caractère de recherche appliquée permettant de répondre aux besoins pratiques des acteurs des différents cas d'étude.

Pour le cas d'étude de la septoriose en Champagne-Ardenne, les principaux partenaires impliqués sont la CRA Champagne-Ardenne, la DRAAF Champagne-Ardenne et Arvalis-Institut du végétal. Mes principaux interlocuteurs sont présentés dans le **Tableau 5.1**.

Tableau 5.1 Principaux interlocuteurs pour le cas d'étude de la septoriose du blé en Champagne-Ardenne

Structure	Représentant	Service
Arvalis	Alexis Decarrier	ingénieur régional délégation Champagne-Ardenne - Céréales à paille (remplacé par Mélanie Franche courant 2015)
Arvalis	Mélanie Franche	ingénieur régional délégation Champagne-Ardenne - Céréales à paille (remplaçante d'Alexis Decarrier)
Arvalis	Charlène Buridant	chargée de l'élaboration BSV-céréales 2014
Arvalis	Amélie Schoonheere	chargée de l'élaboration BSV-céréales 2015
Arvalis	Emmanuelle Gourdain	ingénieur Service Systèmes d'Information et Methodologies. Modèles et outils d'aide à la décision pour les maladies des céréales.
Arvalis	François Piroux	ingénieur Service Systèmes d'Information et Methodologies. Statistiques.
Arvalis	Danièle Simmoneau	responsable des réseaux de surveillance
Arvalis	Gilles Couleaud	ingénieur service Génétique, Physiologie et Protections des Plantes
Arvalis	David Gouache	responsable Service «Valorisations des biotechnologies »
CRA	Vincent Bochu	chargé de mission Écophyto, animateur interfilières du réseau SBT Champagne-Ardenne (remplacé par Karim Benredjem en 2015)
CRA	Karim Benredjem	chargé de mission Écophyto, animateur interfilières du réseau SBT Champagne-Ardenne (remplaçant de Vincent Bochu en 2015)
DRAAF	Gilles Hugerot	responsable du pôle Épidémiologie Maîtrise du Territoire et Expertise Phytosanitaire (jusqu'au 1er septembre 2015)

Le **Tableau 5.2** répertorie les rencontres qui ont permis la co-construction de cet outil. Les échanges ont débuté dès le début de la thèse afin de mieux cerner le contexte de la SBT en région Champagne-Ardenne ainsi que le fonctionnement de l'élaboration de leur BSV. Des informations détaillées sur le protocole d'observation du réseau de SBT et des sorties du modèle épidémiologique SeptoLIS utilisées pour la septoriose ont alors été fournies par les partenaires et discutées collectivement. Les acteurs du BSV nous ont par ailleurs décrit l'organisation et la procédure d'édition du BSV couvrant l'ensemble de la chaîne de traitement de l'information : observations de terrain, saisie dans la base, extraction des données, rédaction et publication du BSV.

Nous avons analysé la procédure d'édition du BSV, et identifié les besoins des partenaires et des pistes d'amélioration. Ainsi nous avons fait trois propositions d'amélioration permettant de mieux valoriser les données issues du réseau de SBT :

- représenter l'évolution de l'incidence à partir des observations brutes
- estimer l'incidence de la maladie chaque semaine à l'échelle régionale et pour chaque groupe de risque (faible, moyen et fort)
- prédire l'incidence de la maladie pour les semaines suivantes à l'échelle régionale et pour chaque groupe de risque (faible, moyen et fort)

Notre objectif était donc d'apporter un outil statistique simple aux acteurs du BSV permettant de répondre à ces besoins.

A noter que des besoins concernant la combinaison des sorties de modélisation avec SeptoLIS avec les données observées en cours de saison avaient également été identifiés. Mais nos travaux n'ont pas permis de répondre à ce point.

Tableau 5.2 : Échanges avec les acteurs du BSV en Champagne-Ardenne ayant permis de co-construire l'outil

Date	Représentants	Lieu	Ordre du jour / Objectifs	Compte-rendu / décisions
30/05/13	Vincent Bochu	email	Prise de contact	Rencontre prévue le 16/06/2013
	Alexis Decarrier		Demande de rencontre pour débiter le travail sur le cas d'étude et aborder le fonctionnement de l'élaboration du BSV	
16/06/13	Charlène Buridant	Arvalis Châlons-en-Champagne	Rencontre avec les acteurs du BSV.	Constats sur le fond du BSV: Analyse clair et simple des observations: moyenne, minimum, maximum des incidences Pas d'analyse statistique sur les données
	Alexis Decarrier		Prise de connaissances de l'historique et le contexte du BSV Analyse du fonctionnement de l'élaboration du BSV Explication du protocole d'observation (Vigicultures®) Explication du modèle épidémiologique (SeptoLIS®): 4 cas types Organisation rédactionnelle (Fond et forme du BSV) Comité relecture et diffusion	Pas d'estimation l'incidence régionale accompagnée d'intervalles de confiance Sous utilisation des données d'observation
	Lucie Michel			Besoins des acteurs du BSV: Regarder l'évolution des observations semaine par semaine Estimer l'incidence moyenne de la maladie semaine par semaine Prédire l'incidence moyenne de la maladie les semaines suivantes accompagnée d'un niveau d'incertitude.
08/07/13	Gilles Couleaud	Arvalis Boigneville	Récupérations des données utilisées pour l'élaboration du BSV	Données Vigicultures® de 2009 à 2014 et sorties du modèle SeptoLIS pour les 4 cas types
	Lucie Michel			
	Danièle Simonneau			
26/03/14	Vincent Bochu	Arvalis Châlons-en-Champagne	Présentation des travaux effectués (premiers graphiques)	Compte-rendu
	François Brun		évolution de la maladie à partir des données brutes	Intérêts de ces graphiques pour améliorer l'élaboration du BSV
	Alexis Decarrier			Evolution annuelle permet de repositionner l'année en cours par rapport aux autres
	David Gouache		évolution de l'incidence médiane pour toute la campagne (1 courbe par année et par feuille) issue du modèle	Décisions
	Emmanuelle Gourdain			Perspective 1: analyser l'effet de réduite la taille du réseau SBT en cours de campagne dans une hypothèse de réduction des coûts
	Gilles Hugerot			Perspective 2: Introduire des covariables dans le modèles (date de semis, variété) afin d'obtenir 3 groupe de risque (faible, moyen, fort)
	David Makowski		prédiction médiane avec intervalle de confiance issue du modèle	Perspective 3: Appliquer la démarche à d'autres cultures/maladies
	Lucie Michel		prédiction par site-année avec intervalle de confiance issues du modèle	Test en cours de campagne des sorties du modèle: envoi des graphiques chaque semaine à Alexis Decarrier et Amélie Schoonheere
	François Piraux			
	Danièle Simonneau			

Chapitre 5 - Co-construction et mise en œuvre opérationnelle de l'outil

19/01/15	Vincent Bochu	Visioconférence (Grignon + Châlon + Boigenville + Toulouse)	Point d'avancement	Pour le graphique par groupe de risque: faire un graphique séparé pour chaque groupe avec comparaison pluri-annuelle comme pour le graphique médian Variabilité inter-site perturbe: seule la tendance suffit Présentation de ces graphiques à prévoir au comité technique puis test en routine à plus large échelle pour la campagne 2015 Perspective: appliquer la démarche pour la septoriose en région Centre
	François Brun		Retour sur ce qui a été envoyé à Alexis Decarrier et Amélie Schoonheere lors de la campagne 2014	
	Alexis Decarrier		Graphiques estimations et prédictions médianes	
	Gilles Hugerot		Graphiques estimations et prédictions par groupe de risque	
	David Makowski		Représentation de la variabilité inter-site (courbe enveloppe)	
	Lucie Michel		Tableau récapitulatif des estimations et prédictions du modèle	
	François Piraux		Transfert méthodologique pour le mildiou de la vigne en région Midi-Pyrénées	
	Danièle Simonneau			
25/02/15	Karim Benredjem	Maison du Bois Châlons en Champagne	Comité technique	Intérêt de la méthode pour traiter d'autres maladies
	François Brun		Présentation du projet SynOEM aux techniciens + résultats pour le cas d'étude septoriose (graphiques)	Variabilité inter-site perturbe: seule la tendance suffit
	Alexis Decarrier			Validation des graphiques: envoi chaque semaine à l'ensemble des acteurs "blé" des graphiques sous forme de rapport optimisé et automatisé
	Gilles Hugerot			Envoi d'un questionnaire d'utilisation de ces rapport en fin de campagne + bilan à prévoir
	Lucie Michel			
	+ comité technique*			
11/09/15	Karim Benredjem	Maison du Bois Châlons en Champagne	Bilan de campagne	Rapports envoyés pendant la campagne 2015 à 31 personnes (10 rapports envoyés en 2015)
	Mélanie Franche		Retour des techniciens sur les rapports de la campagne 2015	Enquête d'utilisation envoyée le 24 juin 2015: 7 retours (23%)
	Lucie Michel			Maintien de l'édition du rapport hebdomadaire pour la campagne 2016 après internalisation de la procédure dans le système informatique d'Arvalis (Boigneville)
	+ comité technique*			Redistribution du rapport par Arvalis -Champagne-Ardenne vers les membres du réseau SBT
21/03/16	Mélanie Franche	Réunion téléphonique	Présentation de l'application web et formation à son utilisation	Utilisation en routine de l'application pour la campagne 2016
	Emmanuelle Gourdain			
*le comité technique étant composé des partenaires de la région pour le réseau SBT:négoces (Digit'Agri, Ets RITARD, COMPAS, SEPAC, EURL Verzeaux), Chambre d'Agriculture (Aube, Ardenne, Marne, Haute-Marne), Terre Inovia, coopératives (SCA EMC2, Acolyance, SCA SCARA, SCA Esternay, Afi Grain, VIVESCIA, Coopérative agricole de juniville,SCA Champagne-Coligny), GRCETA et CETA (CRCETA de l'Aube, GRCETA de Troyes, CETA de Romilly, CETA Champagne, CETA Craie Marne Sud), FREDON Champagne-Ardenne, Agro-industriel (Groupe Soufflet, Cristal Union), association technique (ATPPDA).				

La récupération des observations sous Vigicultures® a été réalisée très rapidement ce qui a permis, après une phase de vérification des données saisies, de comparer différents modèles (Chapitre 2) et de proposer les sorties du modèle qui nous semblait être les plus appropriées dès la 1^{ère} année du projet en mars 2014 (période stratégique afin de mettre en place les premières sorties de modèle en routine pour la campagne 2014). Les premières sorties du modèle sont représentées dans la **Figure 5.1**.

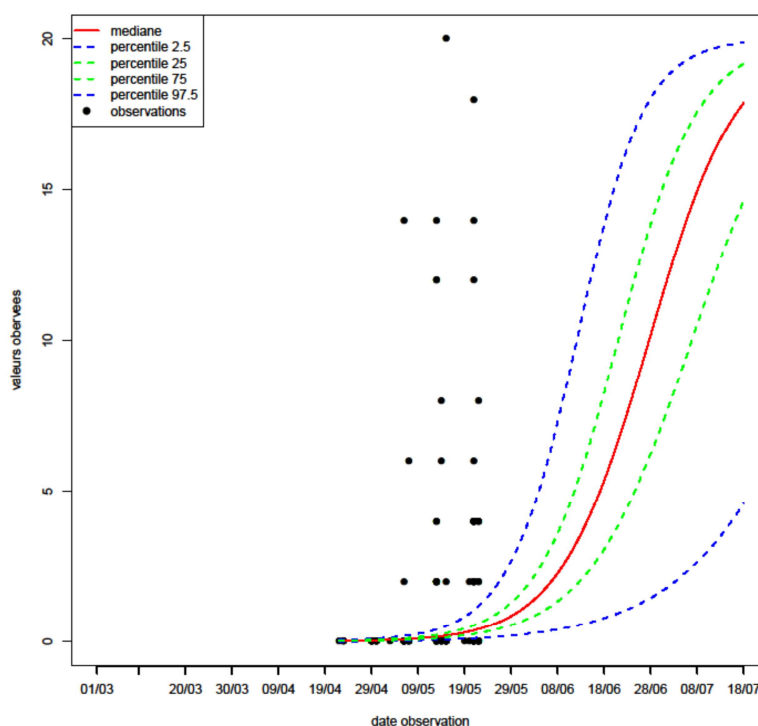


Figure 5.1 : Prédictions médianes pour la feuille 3 au 23 mai 2013. La courbe rouge représente la médiane de l'année 2013, les courbes pointillées bleues et vertes sont les percentiles 2.5, 25, 75 et 97.5. Les ronds noirs sont les observations.

Quelques réajustements ont été faits par rapport au modèle initialement proposé en y ajoutant un effet du groupe de risque permettant de présenter des estimations et prédictions pour chaque groupe de risque. Des améliorations sur la forme du graphique ont été également apportées, en y ajoutant les estimations des années passées, une amélioration de la représentation de l'évolution des données brutes (taille des cercles représentant les observations proportionnelle au nombre de

parcelles observées avec l'incidence considérée) ainsi qu'un intervalle de confiance pour l'estimation de la maladie de l'année en cours.

Les sorties de ce modèle (du type de celles représentées sur les **Figures 5.2** et **5.3**) ont été envoyées chaque semaine aux rédacteurs de la section blé du BSV pendant la campagne 2014. Pour cela, Arvalis me transmettait les observations de la semaine extraites de Vigicultures®, j'ajustais le modèle sur ces données et récupérais les graphiques.

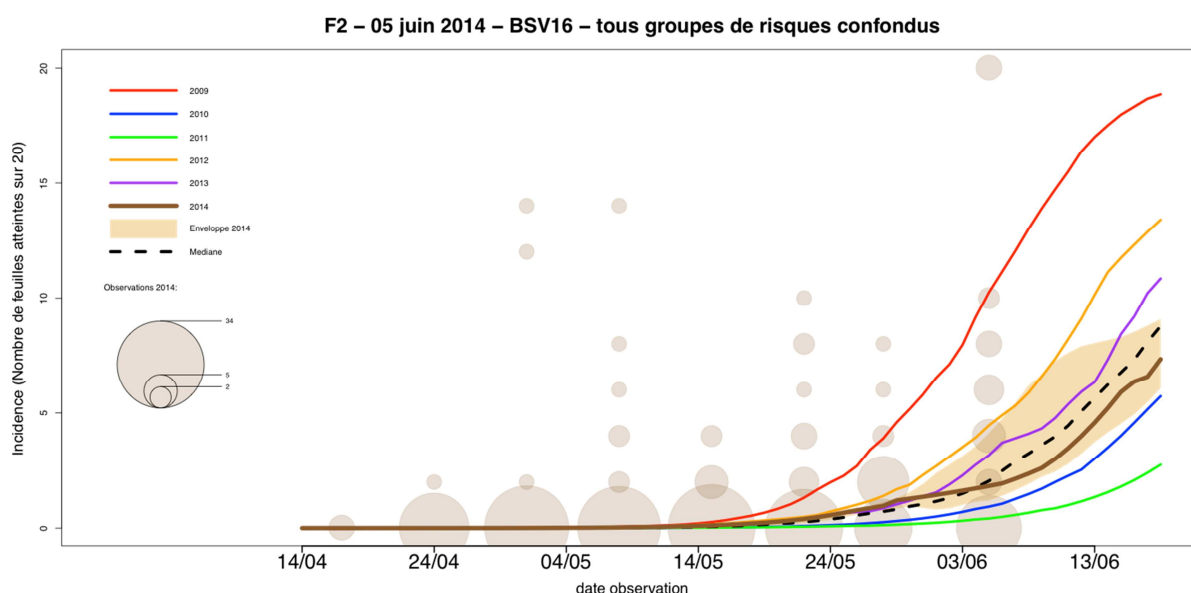


Figure 5.2 : Estimation médiane obtenue en 2014 (courbe marron) et son intervalle de confiance 50% (enveloppe). Les estimations des années antérieures sont représentées en rouge, bleu, vert, orange et violet respectivement pour 2009, 2010, 2011, 2012 et 2013. Les cercles bruns représentent les observations pour la région Champagne-Ardenne en 2014. Leur taille est proportionnelle au nombre de parcelles.

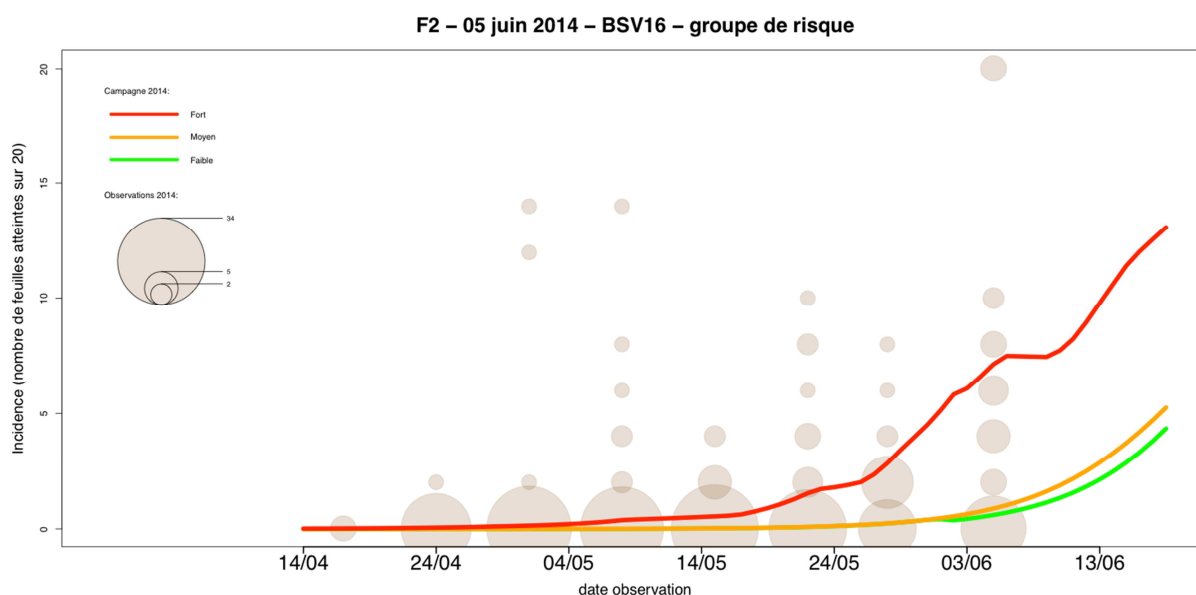


Figure 5.3 : Estimation selon le groupe de risque obtenue en 2014 en rouge pour le groupe de risque fort, en orange pour le groupe de risque moyen et en vert pour le groupe de risque vert. Les cercles bruns représentent les observations pour la région Champagne-Ardenne en 2014. Leur taille est proportionnelle au nombre de parcelles.

Un retour sur ces graphiques a été réalisé début 2015, ainsi quelques modifications ont été effectuées sur la forme des graphiques (volonté d'avoir 4 graphiques séparés : un pour la région et un par groupe de risque ce qui permet de pouvoir se focaliser sur le graphique du groupe de risque auquel appartient la parcelle) et la suppression de la courbe enveloppe représentant la variabilité inter-site, difficilement interprétable d'un premier coup d'œil (enveloppe représentant 50% des sites c'est à dire qu'on a 1 chance sur 2 d'être dans l'enveloppe, 1 chance sur 4 d'être au-dessus et 1 chance sur 4 d'être en-dessous de cette enveloppe). Un test de ces sorties de modèle sous forme d'un rapport automatique, d'une dizaine de pages, a été fait sur la campagne 2015. Il a été envoyé chaque semaine 24h avant la rédaction du BSV après avoir fait l'extraction des observations issues de Vigicultures® (10 rapports sur l'ensemble de la campagne) à l'ensemble des acteurs du BSV blé (31 destinataires). Pour cela, chaque semaine je faisais l'extraction des observations de la semaine dans Vigicultures® (un accès m'a été délivré par Arvalis), j'ajustais le modèle sur ces données et le rapport se créait automatiquement (script développé par Jacques Veslot – CDD ACTA sur le projet SynOEM).

Un exemple de ce rapport est présenté dans la **Figure 5.4** : il se compose d'une en-tête présentant les différents partenaires qui ont collaboré à la co-conception de cet outil et les informations caractérisant la maladie, la culture, la région, la date et la semaine pour lesquels le rapport a été édité. Une table des matières a été ajoutée. Une première partie « Résumé » présente dans un tableau le nombre de parcelles suivies et observations réalisées les feuilles F1, F2, F3 et F4. Les deux dernières colonnes du tableau sont l'estimation régionale de l'incidence de la maladie pour la semaine considérée et la prédiction pour la semaine suivante pour chacune des 3 feuilles (F1, F2 et F3). Une cartographie des parcelles de chaque année selon le groupe de risque est également insérée dans le rapport. La seconde partie concerne la présentation des résultats du modèle pour chaque feuille, de la F1 à la F3 : avec une première sous-partie intitulée « Synthèse » sous forme d'un tableau avec le nombre de parcelles observées à l'échelle régionale et pour chaque groupe de risque depuis le début de la campagne et pour la semaine considérée ainsi que les nouvelles parcelles observées lors de la semaine en cours. Les deux dernières colonnes du tableau de synthèse présentent l'estimation de l'incidence de la maladie pour la semaine considérée et la prédiction pour la semaine suivante. La deuxième sous-partie intitulée « Régional » présente le graphique de la dynamique de l'incidence médiane de la maladie à l'échelle de la région avec en abscisse l'incidence de la maladie (nombre de feuilles atteintes sur 20) et en ordonnée la variable temporelle. La courbe rouge correspond à l'année en cours et chaque autre courbe de couleur représente une année antérieure; la courbe continue rouge représente l'estimation médiane de l'incidence depuis le début de la campagne et celle en tirets représente les prédictions médianes pour les 30 prochains jours. Les ronds gris représentent l'évolution des observations, leur taille est proportionnelle au nombre de parcelles observées qui atteignent la valeur d'incidence située en ordonnée. La 3^{ème} partie pour chaque feuille est le graphique des dynamiques de l'incidence médiane de la maladie pour chaque groupe de risque. Les courbes médianes sont construites uniquement à partir des courbes des parcelles du groupe de risque considéré. La dernière partie du rapport est une notice présentant la règle de décision des groupes de risque ainsi qu'une aide à la lecture des tableaux de synthèse et des graphiques du rapport.



BSV du 27 mai 2015 ¹
Semaine 21
Septoriose du blé d'hiver
Région Champagne-Ardennes

Table des matières

1 Feuille définitive F1	3
1.1 Synthèse	3
1.2 Regional	3
1.3 Par groupe de risque	4
2 Feuille définitive F2	5
2.1 Synthèse	5
2.2 Regional	5
2.3 Par groupe de risque	6
3 Feuille définitive F3	7
3.1 Synthèse	7
3.2 Regional	7
3.3 Par groupe de risque	8
4 Notice	9
4.1 Groupes de risque	9
4.2 Tableau de synthèse	9
4.3 Courbes	10

1. Pour toute question : francois.brun@acta.asso.fr

Résumé

	Parcelles	Observations	Estimation au 27 mai	Prédiction à J+7
F1	50	132	0.0	0.2
F2	49	166	1.7	4.1
F3	50	220	7.9	11.9
F4	51	153		
TOTAL	52	1207		

TABLE 1: Nombre de parcelles et nombre d'observations avec la feuille définitive F, estimation à la date du BSV et prédiction à J+7 de la note régionale de septoriose sur 20 pour la feuille F, pour chacune feuille définitive F1 à F3 (en vert : notes $\leq 1/20$, en orange : notes comprises entre $1/20$ et $4/20$, et en rouge : notes $> 4/20$)

Cartographie des parcelles

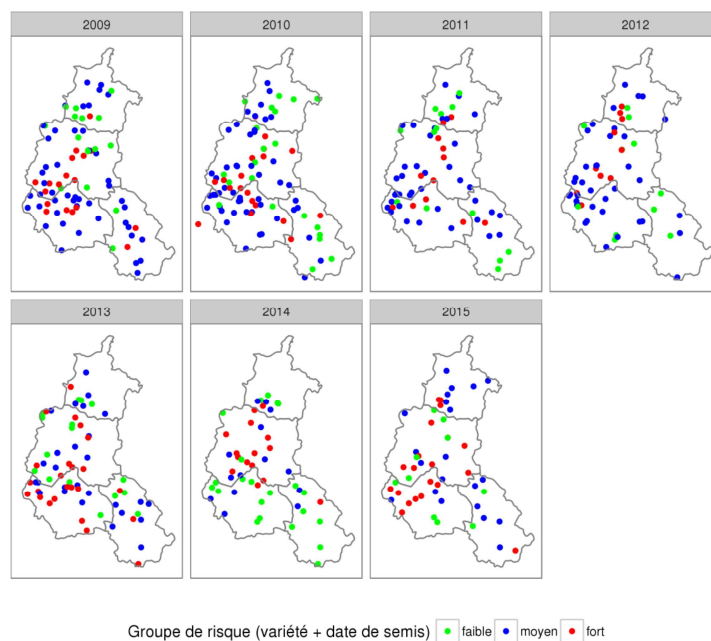


FIGURE 1: Carte des parcelles selon le groupe de risque établi à partir des pratiques culturales (variété, date de semis), par année (en haut à gauche le nombre de parcelles)

1 Feuille définitive F1

1.1 Synthèse

	Nombre de parcelles			Estimation	Prédiction
	Année 2015	Semaine 21	Nouvelles	27 mai	J+7
Regional	52	30	2	0.0	0.2
Faible	10	3	1	0.0	0.1
Moyen	21	13	1	0.0	0.1
Fort	21	14	0	0.2	0.7

TABLE 2: Nombre de parcelles avec feuilles définitives F1 (Cf. notice), estimation au jour du BSV et prédiction à J+7 de la note de septoriose (sur 20)

1.2 Regional

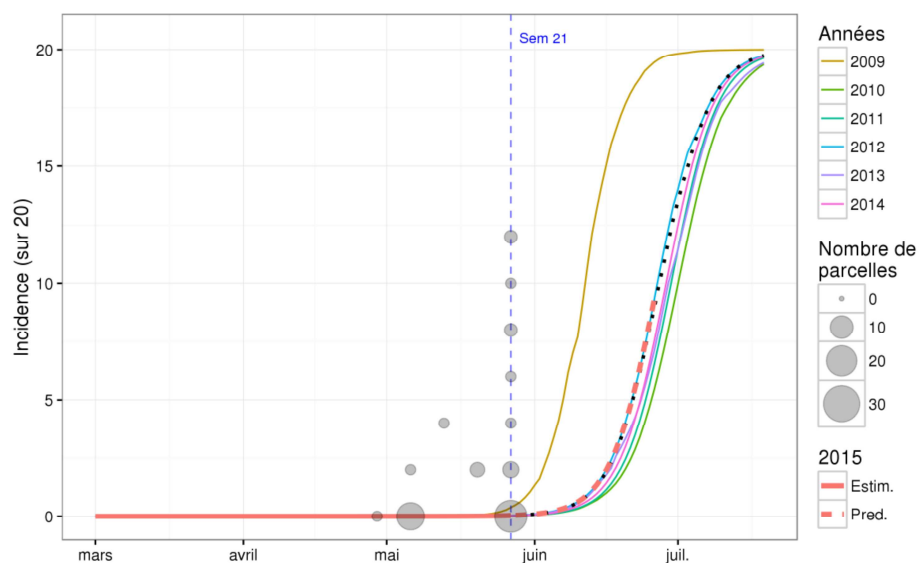


FIGURE 2: Estimations et prédictions de la note de septoriose (sur 20) par année (en rouge : année en cours jusqu'à la date du BSV ; en rouge discontinu : prédiction des 30 jours suivants ; en noir pointillé : la médiane inter-annuelle)

1.3 Par groupe de risque

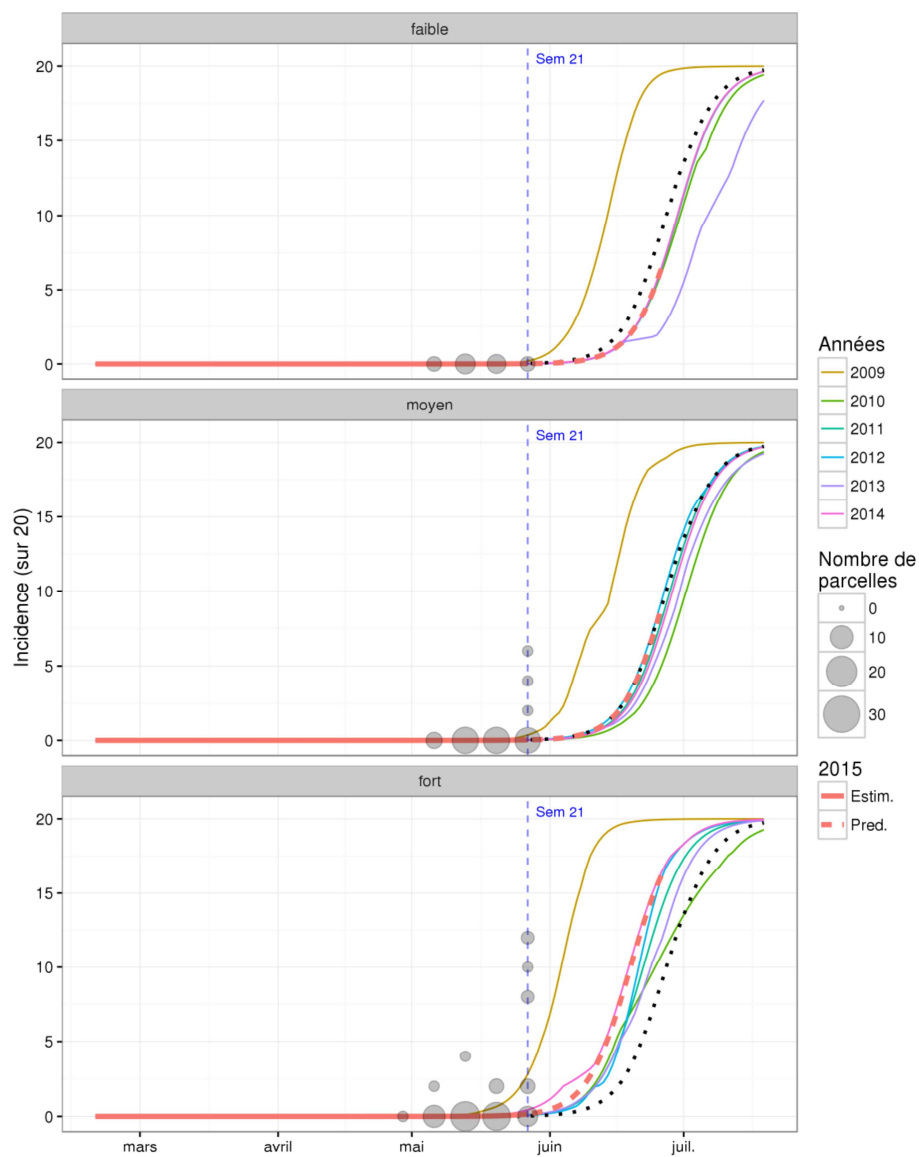


FIGURE 3: Idem préc. - par groupe de risque

2 Feuille définitive F2

2.1 Synthèse

	Nombre de parcelles			Estimation	Prédiction
	Année 2015	Semaine 21	Nouvelles	27 mai	J+7
Regional	52	29	0	1.7	4.1
Faible	10	2	0	0.5	1.5
Moyen	21	13	0	2.4	4.5
Fort	21	14	0	2.4	4.5

TABLE 3: Nombre de parcelles avec feuilles définitives F2 (Cf. notice), estimation au jour du BSV et prédiction à J+7 de la note de septoriose (sur 20)

2.2 Regional

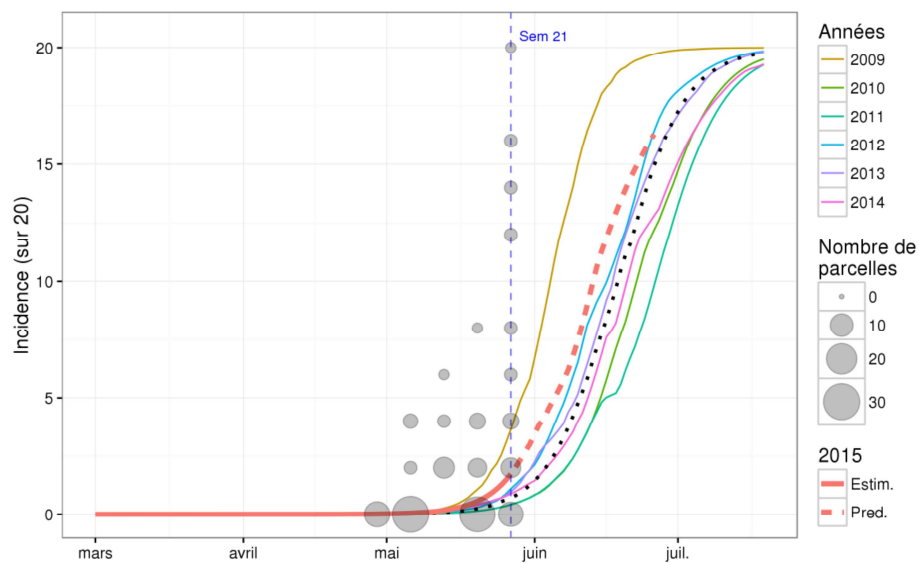


FIGURE 4: Estimations et prédictions de la note de septoriose (sur 20) par année (en rouge : année en cours jusqu'à la date du BSV ; en rouge discontinu : prédiction des 30 jours suivants ; en noir pointillé : la médiane inter-annuelle)

2.3 Par groupe de risque

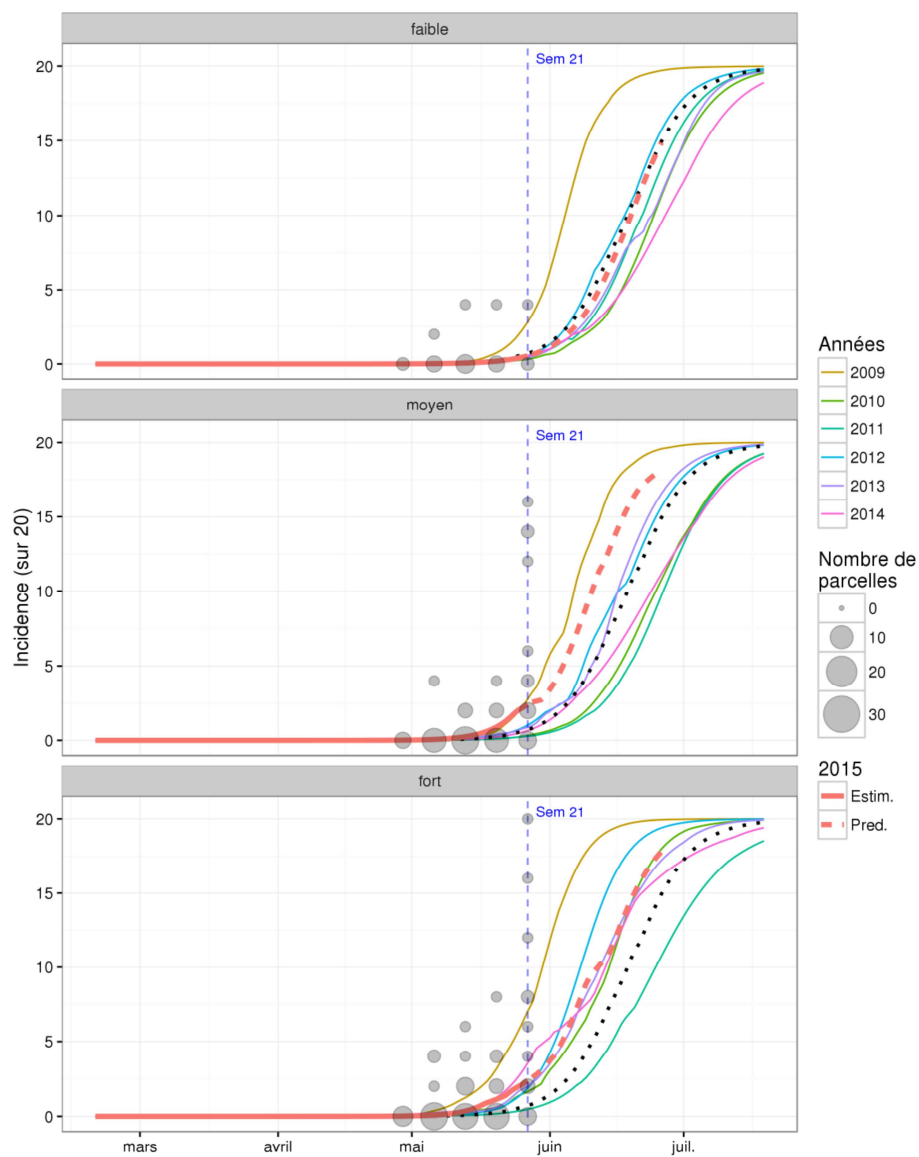


FIGURE 5: Idem préc. - par groupe de risque

3 Feuille définitive F3

3.1 Synthèse

	Nombre de parcelles			Estimation	Prédiction
	Année 2015	Semaine 21	Nouvelles	27 mai	J+7
Regional	52	30	0	7.9	11.9
Faible	10	3	0	5.0	8.7
Moyen	21	13	0	7.6	11.7
Fort	21	14	0	8.9	12.7

TABLE 4: Nombre de parcelles avec feuilles définitives F3 (Cf. notice), estimation au jour du BSV et prédiction à J+7 de la note de septoriose (sur 20)

3.2 Regional

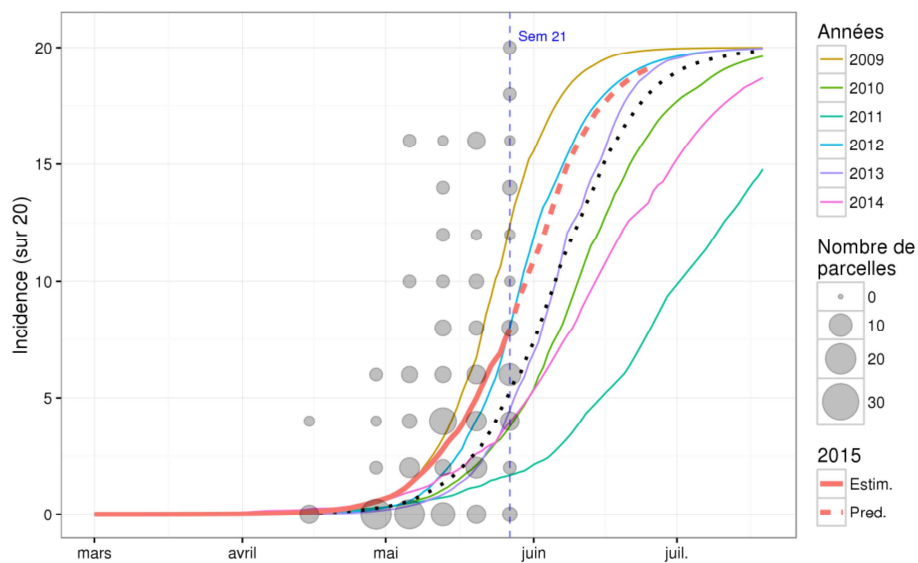


FIGURE 6: Estimations et prédictions de la note de septoriose (sur 20) par année (en rouge : année en cours jusqu'à la date du BSV ; en rouge discontinu : prédiction des 30 jours suivants ; en noir pointillé : la médiane inter-annuelle)

3.3 Par groupe de risque

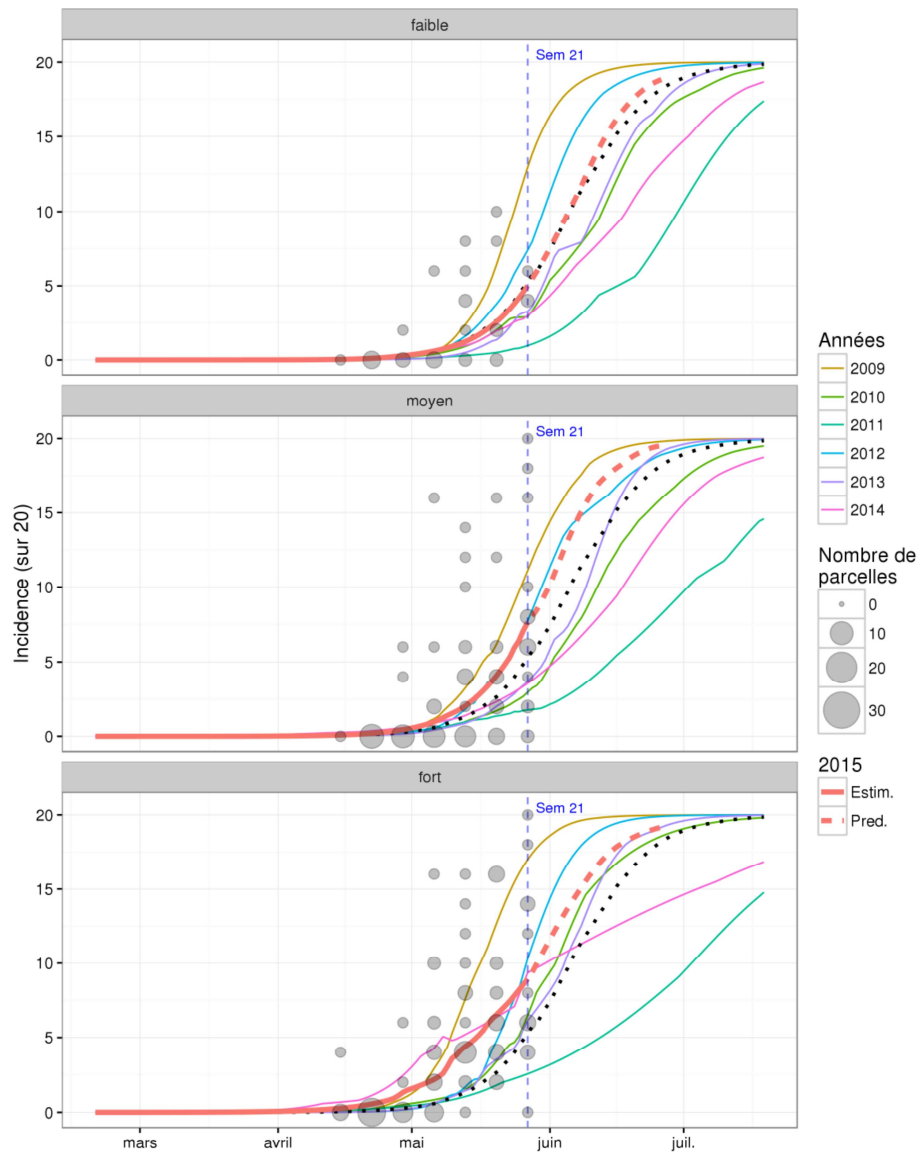


FIGURE 7: Idem préc. - par groupe de risque

4 Notice

4.1 Groupes de risque

Risque	Conditions
Faible	Résistante OU (Intermédiaire ET semis > 25 oct.)
Moyen	(Intermédiaire ET semis < 25 oct.) OU (Sensible ET semis > 25 oct.)
Fort	Sensible ET semis < 25 oct.

TABLE 5: Définition des groupes de risque

Variété	Note
Résistante	6.5 et 7
Intermédiaire	5 à 6
Sensible	4 et 4.5

TABLE 6: Classes de résistance des variétés

4.2 Tableau de synthèse

Descriptif des colonnes :

- **Nombre de parcelles** :
- **Année 2015** : Parcelles de l'année avec au moins une observation de la feuille définitive à la date du BSV ;
- **Semaine 21** : Nombre de parcelles avec au moins une observation de la feuille définitive durant la semaine du BSV ;
- **Nouvelles** : Nombre de parcelles avec une première observation de la feuille définitive dans la semaine du BSV ;
- **Estimation 27 mai** : Estimation de l'incidence (sur 20) à la date du BSV, globale et par groupe de risque, en utilisant la médiane des incidences à la date du BSV estimées par le modèle pour l'ensemble des parcelles déjà observées depuis le début de l'année en cours ;
- **Prédiction à J+7** : Prédiction de l'incidence (sur 20) à J+7 après la date du BSV, prédictions globale et par groupe de risque, en utilisant la médiane des incidences prédites par le modèle pour l'ensemble des parcelles déjà observées depuis le début de l'année.

4.3 Courbes

Descriptif des courbes :

- **Lignes pleines en couleur** : courbes annuelles représentant, pour chaque année, la médiane journalière de l'incidence (sur 20) estimée par la médiane des estimations du modèle pour l'ensemble des parcelles de l'année correspondante ;
- **ligne rouge pleine** : médiane journalière des incidences estimées par le modèle pour l'ensemble des parcelles de l'année en cours jusqu'à la date du BSV ;
ligne rouge discontinue : médiane journalière des incidences prédites par le modèle pour l'ensemble des parcelles de l'année en cours sur les 30 jours qui suivent la date du BSV ;
- **ligne noire pointillée** : médiane journalière des estimations du modèle pour l'ensemble des parcelles de toutes les années.

Sur les **représentations par groupe de risque**, les courbes annuelles et la courbe noire sont obtenues en n'utilisant que les parcelles du groupe correspondant.

Les **points noirs** représentent le nombre d'observations par semaine et par note d'incidence. La taille est proportionnelle au nombre d'observations (d'une semaine donnée avec telle valeur donnée d'incidence), suivant l'échelle à droite du graphe.

Figure 5.4 : Rapport édité le 27 mai 2015 pour l'élaboration du BSV de la semaine 21.

En fin de campagne 2015, une enquête en ligne sur l'utilisation de ce rapport (**Figure 5.5**) a été envoyée aux partenaires de la filière blé (destinataires des rapports durant la campagne 2015) ; nous avons reçu 7 retours de la part d'utilisateurs jouant des rôles différents au sein de l'élaboration du BSV (observateurs, animateur du réseau SBT, rédacteur du BSV, relecture avant publication, responsable filière). L'enquête révèle que les rapports ont été lus entre 1 fois par mois (les relecteurs du BSV étant sollicités pour le comité tournant de relecture 1 fois par mois) et chaque semaine. L'enquête montre un fort intérêt pour notre démarche de valorisation des données. La comparaison faite avec les années antérieures apparaît comme un point important car elle permet de resituer le risque de l'année en cours par rapport aux niveaux de risque passés. Les sorties de notre modèle sont perçues comme complémentaires à celles du modèle épidémiologique SeptoLIS (actuellement prise en compte lors de l'édition du BSV) dans la mesure où elles reflètent la réalité du terrain. Les prédictions d'incidence issues des rapports pourront alimenter le diagnostic de la situation régionale en fonction des groupes de risque au sein du BSV et permettre aux conseillers et agriculteurs d'intégrer ces informations dans leur prise de décisions de stratégie de traitement phytosanitaire.

L'ensemble de la démarche ainsi que le retour des utilisateurs ont été présentés lors de la 11^e conférence internationale sur les maladies des plantes à Tours en décembre 2015 à travers un poster (**Annexe 8**) et un article intitulé « Un outil d'analyse des dynamiques épidémiologiques pour le bulletin de santé du végétal : application à la septoriose du blé » co-écrit avec l'ensemble des partenaires du projet (Michel et al., 2015).

**Questionnaire sur l'utilisation du rapport Septoriose du blé - Campagne 2015-
Champagne-Ardenne**

Chaque semaine un rapport sur la dynamique de l'incidence de la Septoriose du blé en Champagne-Ardenne vous a été envoyé afin de vous donner des informations supplémentaires pour l'élaboration du BSV; merci de répondre à ce court questionnaire anonyme (réponse attendue pour fin juin).

- Rôle en sein de l'élaboration du BSV: (question ouverte)
- Avez-vous lu les rapports? Oui/Non
- Si vous l'avez lu, à quelle fréquence? (question ouverte)
- Si vous ne l'avez pas lu, pourquoi? (question ouverte)
- Si vous avez participé à l'élaboration du BSV, l'avez vous utilisé pour l'élaboration du BSV? Oui/Non
- Si vous avez participé à LA RELECTURE du BSV, l'avez vous utilisé pour la relecture du BSV? Oui/Non
- Si vous l'avez lu, l'avez-vous trouvé pertinent? Oui/Non
- Si vous l'avez lu, qu'est ce que vous aimeriez trouver dedans pour qu'il soit plus pertinent? (question ouverte)

Merci de votre participation

Figure 5.5: Enquête d'utilisation du rapport Septoriose du blé -Campagne 2015- Champagne-Ardenne

5.2 Mise en oeuvre opérationnelle de l'outil

5.2.1 Septoriose du blé

L'automatisation de l'édition du rapport présentée en **Figure 5.4** et la mise en œuvre opérationnelle a été pris en charge par Jacques Veslot (ingénieur en CDD à l'ACTA sur le projet SynOEM), pour que l'édition de ce rapport puisse se faire chaque semaine. Des discussions ont débuté avec Arvalis dès août 2015 pour intégrer l'outil dans leur système informatique afin de faciliter l'utilisation de l'outil par les acteurs régionaux. Cette intégration facilitera le lien direct avec la base de données Vigicultures®, avec le référentiel de sensibilité variétale et l'accès de l'outil par le portail déjà utilisé par les acteurs du BSV. Mais cette internalisation demande de mobiliser différentes compétences ce qui n'a pas pu être réalisé pour la campagne 2016. Aussi, nous avons développé une application web autonome hébergée par l'ACTA permettant aux acteurs d'utiliser facilement et de manière autonome l'outil.

L'application, avec un accès sécurisé, se présente sous la forme d'une interface présentée dans la **Figure 5.6**. Son utilisation comporte 3 étapes :

- Étape 1 : Extraction des données de Vigicultures® depuis 1^{er} janvier de la campagne en cours (cette étape est sécurisée par un compte délivré par Arvalis pour les animateurs régionaux)
- Étape 2 : Importation du fichier d'observations extrait de Vigicultures® (puis cliquer sur « Charger le fichier en cours de saison et Lancer les calculs »).
- Étape 3 : Analyse statistique et récupération des sorties de l'application qui se composent des résultats de l'analyse (**Figure 5.4**) et d'un rapport d'erreur qui recense les erreurs de saisie des observations dans Vigicultures® et indique à l'utilisateur la suppression de certaines observations (celles pour laquelle la date d'observation/ date de semis sont mal saisies, celles dont la variété n'a pas de note de résistance connue, celles qui ont été

malencontreusement traitées, les observations en doublon). Les différents graphiques peuvent être extraits séparément et réutilisé dans le BSV.

ACTA **INRA** **ARVALIS** **AGRICULTURES PRODUISONS AUTREMENT** **ÉCOPHYTO** **ONEMA**

Analyse des dynamiques épidémiologiques pour le BSV - septoriose du blé tendre

Etape 1. Extraire les données de Vigicultures en csv : du 1er janvier à aujourd'hui

Etape 2. Charger le fichier d'extraction de Vigicultures à jour

Choisir le fichier aucun fichier sél.

Charger le fichier en cours de saison et Lancer les calculs

Les calculs sont lancés dès que le fichier est chargé, ainsi que l'édition des rapports. Attention, cette étape d'assimilation des nouvelles données prend du temps (~2min).

le fichier suivant est chargé : ARVALIS_Institut_du_vegetal_export_BTH_2015-01-01_2015-05-27.csv (taille du fichier : 4443715)

Transfert réussi! [Vérifier le fichier sur le serveur](#)

Etape 3. Récupérer le rapport d'analyse

[Le rapport mis à jour](#)

[Fichier de contrôle des données](#)

[Les figures du rapport](#)

[Fichier log d'édition du rapport](#)

Travaux réalisés dans le cadre du projet SynOEM - Mieux profiter de la synergie entre réseaux d'observations, expertise et modélisation pour l'élaboration du Bulletin de Santé du Végétal.

© ACTA – Les Instituts Techniques Agricoles (2013-2016)
<http://www.acta.asso.fr>
149, rue de Bercy 75595 Paris Cedex 12.

Conception et réalisation : François Drun (ACTA), francois.brun@acta.asso.fr, Jacques Veslot (ACTA), Lucie Michel (ACTA/INRA), David Makowski (INRA), Arvalis et Chambre Régionale d'Agriculture

Tous droits de traduction, d'adaptation et de reproduction du site internet (de sa forme et des données) doit faire l'objet d'une demande aux concepteurs. Ce site internet (code PHP, html, base de données sous-jacente, illustration) sont la propriété de l'ACTA et de ses partenaires (notamment pour les illustrations et fiches).

Figure 5.6 : Interface web pour l'édition du rapport d'« Analyse des dynamiques épidémiologiques pour le BSV – septoriose blé tendre ».

Cette application est un prototype mais elle a été présentée aux acteurs en Champagne-Ardenne qui l'utiliseront en routine en cours de campagne et en bilan de campagne en 2016.

En comité régional d'épidémiosurveillance (Réunion du 20 novembre 2015), la région Centre (partenaire du projet synOEM) a également décidé d'utiliser l'application. Nous avons vérifié que le

modèle était pertinent et fiable dans cette région également (Chapitre 3) et ajouté une spécificité de zonage nord-sud à leur demande afin d'avoir des prédictions pour les 2 zones distinctes.

5.2.2 Mildiou de la vigne, cercosporiose de la banane et charançons de la patate douce

Une approche similaire a été réalisée pour le mildiou de la vigne avec l'automatisation d'un rapport pour analyser les données brutes sans approche prédictive dans un premier temps réalisé par François Brun et utilisé en Midi-Pyrénées en 2014 et 2015 (Brun et al., 2015) La version pour la campagne 2016 est en train de voir le jour avec l'intégration du modèle adaptée pour cette maladie (Chapitre3) et une extension possible à la région aquitaine.

Pour la région Guadeloupe, les échanges n'ont eu lieu qu'à partir de mai 2015 et la réflexion est moins aboutie; les acteurs en Guadeloupe sont très intéressés par notre démarche et les sorties du modèles que je leur ai présentés en décembre 2015 et l'ACTA réfléchit à pouvoir intégrer ce cas d'étude dans un nouveau projet dans le but de poursuivre la collaboration et d'apporter un outil sur ces cultures.

Conclusion générale

La surveillance biologique obligatoire du territoire et l'édition du Bulletin de Santé du Végétal ont été mises en place en 2009, à la suite de l'adoption du plan Écophyto. Ces deux mesures ont vocation à aider les agriculteurs et leurs conseillers à raisonner leur stratégie de traitements phytosanitaires. Mes travaux ont conduit au développement d'une méthode d'analyse de données permettant aux acteurs du BSV de traiter les informations dont ils disposent chaque semaine pour élaborer le BSV. Cette méthode permet d'améliorer la valorisation des observations issues du réseau d'épidémiosurveillance en cours de saison et en fin de campagne.

La méthode proposée est basée sur un modèle statistique permettant d'estimer l'incidence d'une maladie ou l'abondance d'un ravageur dans un réseau de parcelles agricoles. Ce modèle statistique est de type mixte linéaire généralisé. Il peut être facilement ajusté aux données collectées dans les réseaux d'épidémiosurveillance en tenant compte de certaines caractéristiques des parcelles (ex : date de semis et sensibilité variétale). Cette approche a été évaluée de manière approfondie sur la septoriose du blé, puis sur plusieurs autres bioagresseurs. Dans le cas de la septoriose, le modèle permet d'estimer, chaque semaine, l'incidence de la maladie en cours de saison et de prédire pour les semaines à venir l'évolution de cette incidence à la fois à l'échelle régionale, à l'échelle de groupes de parcelles (correspondant à des niveaux de risque contrastés), et à l'échelle des parcelles individuelles.

L'unique contrainte pour pouvoir ajuster le modèle est la disponibilité d'observations d'incidence ou d'abondance. La collecte de telles observations est obligatoire depuis le plan Écophyto, mais la taille du réseau d'épidémiosurveillance influence la qualité d'estimation du modèle. Un des chapitres de cette thèse a ainsi été consacré à l'analyse de la relation entre la taille du réseau et la précision du modèle. Les résultats obtenus m'ont permis de déterminer les conséquences d'une diminution de la taille du réseau à la fois sur les performances du modèle et sur les coûts financiers associés au suivi du réseau de surveillance. Dans le cas de septoriose en Champagne-Ardenne, j'ai montré qu'il était possible de réduire la taille du réseau de surveillance en cours de saison en appliquant une stratégie de stratification des parcelles selon leur incidence et d'échantillonner aléatoirement dans chacune des strates déterminées pour obtenir un plus petit échantillon. Cette approche permet de réduire la taille de l'échantillon de parcelles sans trop diminuer la précision des estimations d'incidence.

A la fin de ma thèse, le modèle développé pour la septoriose a été intégré dans un outil informatique opérationnel co-construit avec l'ensemble des acteurs du BSV en région Champagne-Ardenne. Cet outil génère automatiquement un rapport automatique, délivré via une interface utilisateur disponible sur internet. Chaque semaine, il reprend synthétiquement les informations issues du réseau d'épidémiosurveillance sous forme de tableaux ce qui permet d'avoir rapidement l'ensemble des informations descriptives nécessaires à l'élaboration du BSV. Les graphiques inclus dans le rapport présentent (i) l'évolution des observations du réseau, (ii) l'estimation de l'incidence à l'échelle régionale et du groupe de risque depuis le début de la campagne, (iii) les prédictions de l'incidence pour les semaines à venir et (iv) permettent de comparer l'année en cours aux années antérieures. Ce rapport est disponible chaque semaine pour aider les acteurs à élaborer le BSV. Il est utilisé au cours de la saison en 2016 pour l'élaboration des BSV et il permettra d'établir un bilan de campagne pour la région Champagne-Ardenne.

J'ai montré que le modèle initialement développé pour la septoriose pouvait facilement s'adapter à d'autres régions, d'autres cultures et maladies. Le modèle proposé a ainsi été adapté pour analyser la dynamique de la septoriose du blé en région Centre. Il est déjà opérationnel pour l'édition du BSV dans cette région. Le modèle a également été adapté pour le mildiou de la vigne en région Midi-Pyrénées. Dans ce cas, le modèle permet d'estimer la probabilité qu'un cep de vigne soit atteint par le mildiou. En région Guadeloupe, le modèle a été modifié pour analyser les observations de cercosporiose jaune du bananier et estimer l'évolution de la probabilité de dépasser le seuil d'État d'Évolution 250 (passage d'un niveau de contamination faible à moyen) en cours de saison. Enfin le modèle a été également adapté pour analyser les données de comptage de charançons capturés dans des cultures de patate douce en Guadeloupe. A terme, un outil opérationnel d'édition du BSV pourrait être développé en Guadeloupe pour suivre l'évolution des principaux bioagresseurs de cette région.

Ces résultats montrent que notre démarche méthodologique a suscité un vif intérêt de la part des acteurs du BSV. Ce succès est due à plusieurs facteurs : l'implication précoce des acteurs du BSV dans le travail, la disponibilité des données épidémiologiques du réseau de surveillance biologique du territoire, et l'utilisation d'outils statistiques robustes permettant une analyse rapide des

données. A l'avenir, cette démarche pourrait être étendue pour intégrer les connaissances des experts régionaux, en plus des données épidémiologiques.

Références bibliographiques

ACTA (Association de coordination technique agricole), 2013. Projet SynOEM: Mieux profiter de la synergie entre réseaux d'observation, expertise et modélisation pour l'élaboration du Bulletin de Santé du Végétal. En réponse à l'appel à projet Pour et Sur le Plan Écophyto (PSPE) sur les trois dispositifs du Plan: Epidémiosurveillance/DEPHY/Indicateurs. Document du projet, 21p. Disponible sur : <http://www.modelia.org/moodle/course/view.php?id=55> Consulté le: 04/004/2016.

Agresti, A., 2002. Categorical data analysis, second ed. New York: John Wiley & Sons.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Ansar, M., Cheema, N.M., and Leitch, M.H. (2010). Effect of agronomic practices on the development of Septoria leaf blotch and its subsequent effect on growth and yield components of wheat. *Pakistan Journal of Botany*, 43(3), 2125–2138.

Apel, H., Paudyal, M.S., and Richter, O., 2003. Evaluation of treatment strategies of the late blight *Phytophthora infestans* in Nepal by population dynamics modelling. *Environmental Modelling & Software*, 18(4), 355–364. doi:10.1016/S1364-8152(02)00106-8

Arvalis - Institut du végétal, 2015. Actualités phytosanitaires. Choisir & Décider - Synthèse nationale, 2014-2015, 6–17.

Aubertot, J.N., Barbier, J.M., Carpentier, A., Gril, J.J., Guichard, L., Lucas, P., Savary, S., and Voltz M., 2007. Pesticides, agriculture et environnement. Réduire l'utilisation des pesticides et en limiter les impacts environnementaux. Congrès : Expertise Scientifique INRA-CEMAGREF (2005/12; Paris, FRA). Ed. Quae, Versailles. 119p.

Audsley, E., Milne, A., and Paveley, N., 2005. A foliar disease model for use in wheat disease management decision support systems. *Annals of Applied Biology*, 147(2), 161–172. doi:10.1111/j.1744-7348.2005.00023.x.

Barbetti, M.J., Banga, S.S., and Salisbury, P.A., 2012. Challenges for crop production and management from pathogen biodiversity and diseases under current and future climate scenarios – Case study with oilseed Brassicas. *Field Crops Research*, 127, 225–240. doi:10.1016/j.fcr.2011.11.021.

Bernard, J.-L., 2007. Protection chimique des plantes cultivées et durabilité. *Oilseeds and fats, Crops and Lipids*, 14(6), 332-344.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., and White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. doi:10.1016/j.tree.2008.10.008

Brun, F., Veslot, J., Michel, L., Cichosz, B., Petit, A., and Makowski, D., 2015. Quelles pistes d'amélioration pour mieux valoriser les données et les simulations dans le bulletin de sante du végétal ? *AFPP. 11ème conférence Internationale sur les maladies et les plantes – Tours – 7 et 8 Décembre 2015.*

Champoiseau. P., Daugrois. J.H., Grossard. F., Guyader. S., Hery. M., Lombion. K., and Osseux. J., 2011. BSV n°2 Juillet 2011 Région Guadeloupe Filières de Diversification Végétal.

CRAC (Chambre Régionale d'Agriculture du Centre), 2014. Bulletin de santé du végétal – Céréales à paille – 27/ 05/2014- n°22- Région Centre. Disponible sur: http://www.centre.chambagri.fr/fileadmin/documents/CRA_Centre/Developpement/Bulletin_Sant_e_Vegetal/Grandes_cultures/Cereales_a_paille/2014-2015/BSV_cereales_paille_22..pdf Consulté le: 04/04/2016.

CRACA (Chambre Régionale d'Agriculture de Champagne-Ardenne), 2012 Bulletin de santé du végétal – Grandes Cultures – Champagne-Ardenne - 24/05/2012. Disponible sur: http://www.champagric.fr/fileadmin/documents_craca/BSV/BSV_2012/14_-_BSV_Grandes_cultures_S21.pdf . Consulté le: 04/04/2016.

Decarrier, A., and Gagliardi, E., 2015. Stratégies régionales fongicides. Normandie, Picardie, Nord-Pas-de-Calais, Champagne-Ardenne. Contre les résistances à la septoriose. *Perspectives agricoles*, (418), 17.

Delph, B.R., Stowell, L.J., and Marois, J.J., 1986. Evaluation of fields sampling techniques for estimation of disease incidence. *The American Phytopathological Society*, 76(12), 1299-1305.

Denon, D., and Mauléon, H., 2004 - Le charançon de la patate douce en Guadeloupe, *Cylas formicarius*, menace gravement la survie de la culture. *Phytoma* n°573, 14-15. ☐

DGAL-SDQPV (Direction générale de l'alimentation. Sous-direction de la qualité et de la protection des végétaux), 2014. Rapport Annuel de Surveillance Biologique du Territoire. Disponible sur : <http://agriculture.gouv.fr/rapport-annuel-de-surveillance-biologique-du-territoire> Consulté le: 4 avril 2016.

Disthaporn, S., Hau, B., and Kranz, J., 1993. Comparison of sampling procedures for two rice diseases : leaf blast tungro. *Plant Pathology*, 42, 313-323.

DRAAF-CA (Direction régionale de l'alimentation, de l'Agriculture et de la forêt – Champagne-Ardenne), 2014. Agreste Champagne-Ardenne 2013: une campagne en demi-teinte. Service régionale de l'information statistique et économique.

DRAAF-SRAL (Direction régionale de l'alimentation, de l'Agriculture et de la forêt - Service régional de l'alimentation), 2010. Bulletin d'information phytosanitaire mensuel n°3 – 25/03/2010. Préfecture de Champagne-Ardenne DRAAF – Service régional de l'alimentation. Disponible sur : <http://www.champagne-info.net/avert/23.pdf> Consulté le: 04/04/2016.

- El Chartouni, L., Tisserant, B., Siah, A., Duyme, F., Durand, R., Halama, P., and Reignault, P.,** 2012. Evolution of *Mycosphaerella graminicola* at the wheat leaf and field levels. *Phytopathologia Mediterranea*, 51(2), 332–339.
- El Jarroudi, M., Delfosse, P., Maraite, H., Hoffmann, L., and Tychon, B.,** 2009. Assessing the accuracy of simulation model for Septoria leaf blotch disease progress on winter wheat. *Plant Disease*, 93(10), 983–992.
- Eyal, Z., Scharen, A.L., Prescott, J.M. , and van Ginkel, M.,** 1987. The Septoria Diseases of Wheat: Concepts and methods of disease management. Mexico, D.F.: CIMMYT.
- Fones, H. and Gurr, S.,** 2015. The impact of Septoria tritici Blotch disease on wheat: An EU perspective. *Fungal Genetics and Biology*, 79, 3–7. doi:10.1016/j.fgb.2015.04.004.
- Fong, Y., Rue, H., and Wakefield, J.,** 2010. Bayesian inference for generalized linear mixed models. *Biostatistics*, 11(3), 397–412. doi:10.1093/biostatistics/kxp053
- Fouré, E.,** 1994. Leaf spot diseases of banana and plantain caused by *Mycosphaerella musicola* and *M. fijiensis*. In: DR Jones (ed) The Improvement and Testing of Musa: A Global Partnership. Proceedings of the First Global Conference of the International Musa Testing Program Held at FHIA, Honduras, April 27–30, 1994.
- Ganry, J., de Lapeyre de Bellaire, L., and Mourichon, X.,** 2008. A biological forecasting system to control Sigatoka disease of bananas and plantains. *Fruits*, 63(6), 381–387. doi:10.1051/fruits:2008038
- Gelman, A. and Rubin, D.B.,** 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511.

GEVES. (2014). Catalogue Officiel. Disponible sur : http://cat.geves.info/Cat/Cat_Pag_Acceuil/FA8AAKT1jodhSnFoUWZyaVZYDAA?WD_ACTION=MENU&ID=M29 Consulté le: 04/04/2016.

Gessler, C., Pertot, I., and Perazzolli, M., 2011. Plasmopara viticola: a review of knowledge on downy mildew of grapevine and effective disease management. *Phytopathologia Mediterranea* 50, 3–44

Gouache, D. and Couleaud, G., 2010. Fongicides céréales. Positionnement des traitements ciblant la septoriose: un enjeu de 5q/ha. *Perspectives agricoles*, (365), 42–45.

Gouache, D., Bensadoun, A., Brun, F., Pagé, C., Makowski, D., and Wallach, D., 2012. Modelling climate change impact on Septoria tritici blotch (STB) in France: Accounting for climate model and disease model uncertainty. *Agricultural and Forest Meteorology*, 170, 242–252. doi:10.1016/j.agrformet.2012.04.019.

Hardwick, N.V., Jones, D.R., and Slough, J.E., 2001. Factors affecting diseases of winter wheat in England and Wales, 1989–98. *Plant Pathology*, 50(4), 453–462.

Henze, M., Beyer, M., Klink, H., and Verreet, J.-A., 2007. Characterizing meteorological scenarios favorable for Septoria tritici infections in wheat and estimation of latent periods. *Plant Disease*, 91(11), 1445–1449.

Hughes, G., and Gottwald, T.R., 1998. Survey methods for assessment of citrus tristeza virus incidence. *Phytopathology* 88 (7), 715–723.

IFV (Institut Français de la vigne et du vin), 2010. Protocole de suivi d'une parcelle témoin dans le cadre de la Sureveillance Biologique du Territoire pour la rédaction du BSV en Aquitaine.

INSERM (Institut nationale de la santé et de la recherche médicale), 2013. Pesticides. Effets sur la santé. Collection expertise collective, Paris. Disponible sur :

http://www.ipubli.inserm.fr/bitstream/handle/10608/4820/Chapitre_1.html Consulté le: 04/04/2016.

Jermini, M., Blaise, P., and Gessler, C., 2010. Quantitative effect of leaf damage caused by downy mildew (*Plasmopara viticola*) on growth and yield quality of grapevine 'Merlot' (*Vitis vinifera*). *VITIS-Journal of Grapevine Research*, 49(2), 77.

Jørgensen, L.N., Hovmøller, M.S., Hansen, J.G., Lassen, P., Clark, B., Bayles, R., et al., 2014. IPM strategies and their dilemmas including an introduction to www.eurowheat.org. *Journal of Integrative Agriculture*, 13(2), 265–281. doi:10.1016/S2095-3119(13)60646-2.

Jullien, J., 2013. Surveillance phytosanitaire : des Avertissements Agricoles aux Bulletins de santé du végétal. *Jardin de France* n°621, Dossier Jardinage, alliance de la passion et de l'engagement. Janvier-Février 2013. 9p.

Juroszek, P. and Von Tiedemann, A., 2015. Linking plant disease models to climate change scenarios to project future risks of crop diseases: a review. *Journal Plant Disease Protect*, 122, 3–15.

King, J.E., Cook, R.J., and Melville, S.C., 1983. A review of Septoria diseases of wheat and barley. *Annals of Applied Biology*, 103, Issue 2, 345–373.

Kim, K.-H., Cho, J., Lee, Y.H., and Lee, W.-S., 2015. Predicting potential epidemics of rice leaf blast and sheath blight in South Korea under the RCP 4.5 and RCP 8.5 climate change scenarios using a rice disease epidemiology model, EPIRICE. *Agricultural and Forest Meteorology*, 203, 191– 207. doi:10.1016/j.agrformet.2015.01.011.

Lassoudière, A., 2007. Le bananier et sa culture. Ed. Quae, Versailles. 383p.

Lhoste, J., and Grison, P., 1989. La phytopharmacie française. Chronique historique. Versailles, INRA, 279p.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D., 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337.

Luo, W., Pietravalle, S., Parnell, S., van den Bosch, F., Gottwald, T.R., Irely, M.S., and Parker, S.R., 2012. An improved regulatory sampling method for mapping and representing plant disease from a limited number of samples. *Epidemics* 4, 68–77. doi:10.1016/j.epidem.2012.02.001

MAAF (Ministère Français en charge de l'agriculture), 2008. Plan Écophyto 2018. (2008, September) Disponible sur : <http://agriculture.gouv.fr/alimentation/ecophyto-2018-un-plan-pour-reduire-de-moitie-lusage-des-pesticides> Consulté le: 04/04/2016.

MAAF (Ministère Français en charge de l'agriculture), 2015a. Plan Écophyto II. (2015, October). Disponible sur: http://agriculture.gouv.fr/sites/minagri/files/151022_ecophyto.pdf Consulté le: 04/04/2016.

MAAF (Ministère Français en charge de l'agriculture), 2015b. Le BSV, pour quoi ? Pour qui ? – 10/12/2015. Alim'agri : Site du ministère de l'agriculture, de l'agroalimentaire et de la forêt. Disponible sur : <http://agriculture.gouv.fr/le-bsv-pour-quoi-pour-qui>. Consulté le: 04/04/2016.

MAAF (Ministère Français en charge de l'agriculture), 2015c. BSV : Comment les données sont-elles centralisées ? – 17/12/2015. Alim'agri : Site du ministère de l'agriculture, de l'agroalimentaire et de la forêt. Disponible sur : <http://agriculture.gouv.fr/bsv-comment-les-donnees-sont-elles-centralisees>. Consulté le: 04/04/2016.

MAAF (Ministère Français en charge de l'agriculture), 2016. Bulletins de santé du végétal – 02/02/2016. Alim'agri : Site du ministère de l'agriculture, de l'agroalimentaire et de la forêt. Disponible sur : <http://agriculture.gouv.fr/bulletins-de-sante-du-vegetal>. Consulté le: 04/04/2016.

Madden, L.V., Turechek, W.W., and Nita, M., 2002. Evaluation of generalized linear mixed models for analyzing disease incidence data obtained in designed experiments. *Plant disease*, 86(3), 316–325.

Makowski, D., Vicent, A., Pautasso, M., Stancanelli, G., and Rafoss, T., 2014. Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*. *European Journal of Plant Pathology*, 139(1), 79–94. doi:10.1007/s10658-013-0365-6.

Mauléon, H., and Denon, D., 2008. Etude de la répartition en Guadeloupe de *Cylas formicarius*, charançon de la patate douce, et essai d'une méthode de lutte biologique. Rapport final - INRA Centre Antilles-Guyane Unité de Recherche en Productions Végétales.

Meredith, D.S., 1970. Banana Leaf Spot Disease (Sigatoka) caused by *Mycosphaerella muscol* leach. *Phytopathological papers*, No. 11, Commonwealth Mycological Institute, Kew, Surrey, UK, 147p.

Michel, L., Decarrier, A., Franche, M., Bochu, V., Benredjem, K., Hugerot, G., Simonneau, D., Piraux, F., Gourdain, E., Veslot, J., Makowski, D., and Brun, F., 2015. Un outil d'analyse des dynamiques épidémiologiques pour le bulletin de santé du végétal: application a la septoriose du blé. *AFPP. 11ème conférence Internationale sur les maladies et les plantes – Tours – 7 et 8 Décembre 2015.*

Mila, A.L. and Carriquiry, A.L., 2004. Bayesian analysis in plant pathology. *Phytopathology*, 94(9), 1027–1030.

Mila, A.L. and Ngugi, H.K., 2011. A Bayesian approach to meta-analysis of plant pathology studies. *Phytopathology*, 101(1), 42–51.

- Moreau, J.-M. and Maraite, H.,** 1999. Integration of knowledge on wheat phenology and *Septoria tritici* epidemiology into a disease risk simulation model validated in Belgium. *Aspects of Applied Biology*, 55, 1–6.
- Moreau, J.-M. and Maraite, H.,** 2000. Development of an interactive decision-support system on a Web site for control of *Mycosphaerella graminicola* in winter wheat. *EPPO Bulletin*, 30(1), 161–163.
- Ojiambo, P.S., and Scherm, H.,** 2010. Efficiency of adaptive cluster sampling for estimating plant disease incidence. *Phytopathology*, 100 (7), 663–670.
- Parsons, D. and Te Beest, D.** (2004). Optimising fungicide applications on Winter Wheat using genetic algorithms. *Biosystems Engineering*, 88(4), 401–410. doi:10.1016/j.biosystemseng.2004.04.012.
- Penaud, A., Printz, P., Maumené, C., Couleaud, G., Walker, A. S., Gouache, D., Ney, B., and Pinochet, X.,** 2011. Des outils de raisonnement pour limiter et rendre durable la lutte chimique contre les bioagresseurs en Grandes Cultures. *Innovations Agronomiques*, 17, 1–14.
- Pérez, M.L. del P., Isas, M.G., Salvatore, A.R., Gastaminza, G., and Trumper, E.V.,** 2015. Optimizing a fixed-precision sequential sampling plan for adult *Acrotomopus atropunctellus* (Boheman) (Coleoptera: Curculionidae), new pest on sugarcane. *Crop Protection* 74, 9–17. doi:10.1016/j.cropro.2015.03.024
- Philibert, A., Loyce, C., and Makowski, D.,** 2014. Predicting nitrous oxide emissions with a random-effects model. *Environmental Modelling and Software*, 61, 12–18.
- Piepho, H.P.,** 1999. Analysing disease incidence data from designed experiments by generalized linear mixed models. *Plant pathology*, 48(5), 668–674.

Pietravalle, S., Shaw, M. W., Parker, S. R., and Van Den Bosch, F., 2003. Modeling of relationships between weather and *Septoria tritici* epidemics on winter wheat: a critical approach. *Phytopathology*, 93(10), 1329–1339.

Poitout, S.-H., and Leclant, F., 1986. Progrès en protection des plantes cultivées et évolution sociale et industrielle aux XIX^e et XX^e siècles. Culture technique n°16 -Technologies agro-alimentaires.

Polley, R.W. and Thomas, M.R., 1991. Surveys of diseases of winter wheat in England and Wales, 1976–1988. *Annals of Applied Biology*, 119(1), 1–20.

Ponomarenko, A., Goodwin, S.B., and Kema, G.H.J., 2011. *Septoria tritici* blotch (STB) of wheat. *Plant Health Instructor*. doi:10.1094/PHI-I-2011-0407-01

Potier, D., 2014. Pesticides et agro-écologie – Les champs du possible. Rapport de Dominique Potier, député de meurthe-et-moselle, au premier ministre Manuel Valls .

Pouget, R., 1990. Histoire de la lutte contre le phylloxera de la vigne en France (1968-1895). Ed. INRA, Paris. 158p.

Prado, R., and West, M., 2010. Time Series: Modeling, Computation, and Inference, 1st ed. Chapman and Hall/CRC.

R Development Core Team, 2014 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07- 0, URL <http://www.r-project.org>.

Regnault-Roger, C., 2005. Enjeux phytosanitaires pour l’agriculture et l’environnement. Londre-Paris- New-York: TEC et DOC, 1013p

Regnault-Roger, C., 2014. Produits de Protection des Plantes. Innovation et sécurité pour une agriculture durable. Paris: TEC et DOC, 353p.

Robert, C., Fournier, C., Andrieu, B., and Ney, B., 2008. Coupling a 3D virtual wheat (*Triticum aestivum*) plant model with a *Septoria tritici* epidemic model (Septo3D): a new approach to investigate plant–pathogen interactions linked to canopy architecture. *Functional Plant Biology*, 35(10), 997. doi:10.1071/FP08066.

Rocha, H.S., Pozza, E.A., Uchôa, C.D.N., Cordeiro, Z.J.M., De Souza, P.E., Sussel, Ân.A.B., and De Rezende, C.A., 2012. Temporal Progress of Yellow Sigatoka and Aerobiology of *Mycosphaerella musicola* Spores: Temporal Progress of Yellow Sigatoka and *M. musicola* Spores. *Journal of Phytopathology*, 160(6), 277–285. doi:10.1111/j.1439-0434.2012.01897.x

Rossi, V., Caffi, T., and Salinari, F., 2012. Helping farmers face the increasing complexity of decision-making for crop protection. *Phytopathologia Mediterranea*, 51(3), 457–479.

Savary, S., Stetkiewicz, S., Brun, F., and Willocquet, L., 2015. Modelling and mapping potential epidemics of wheat diseases—examples on leaf rust and *Septoria tritici* blotch using EPIWHEAT. *European Journal of Plant Pathology*, 142(4), 771–790. doi:10.1007/s10658-015-0650-7.

Sigvald, R., 2012. Risk Assessments for Pests and Diseases of Field Crops, especially Forecasting and Warning Systems. In: Sustainable Agriculture / [ed] Christine Jakobsson, Uppsala: Baltic University Press ,185-201.

Sim IV, T., Willis, W.G., and Eversmeyer, M.G., 1998. Kansas Plant Disease Survey. *Plant Disease*, 72(10), 832–836.

Simonneau, D., Couleaud, G., Flodrops, Y., Bodilis, A.M., and Duyme, F., 2009. Vigicultures® Moyens de validation des observations. 5p (unpublished report)

Simonneau, D., Chollet, D., Bouveris, N., and Brochard, M., 2010. Surveillance Vigicultures® adopté par 17 régions. *Perspectives agricoles*, 372, 52–54.

Simonneau, D., Duyme, F., Chollet, D., Perrollet, S., Bouveris, N., and Brochard, M., 2011a. Une méthodologie de calibration de réseaux d'épidémiosurveillance. AFPP. 4ème conférence Internationale sur les méthodes alternatives en protection des cultures – Lille – 8, 9 et 10 Mars 2011.

Simonneau, D., Taupin, P., Couleaud Gilles, Maufra, J.-Y., and Robin, N., 2011b. Vigicultures® Mode opératoire observation Blés hiver. 25p

Sine, M., Morin, E., Simonneau, D., Brochard, M., De Cosnac, G., and Escriou, H., 2010. VIGICULTURES–An early warning system for crop pest management. In Scientific and Technical Information and Rural Development IAALD XIIIth World Congress, Montpellier (pp. 26–29).

Small, I. M., Joseph, L., and Fry, W.E., 2015. Development and implementation of the BlightPro decision support system for potato and tomato late blight management. *Computers and Electronics in Agriculture*, 115, 57–65. doi:10.1016/j.compag.2015.05.010.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 64(4), 583–639.

Suffert, F., Sache, I., and Lannou, C., 2011. Early stages of Septoria tritici blotch epidemics of winter wheat: build-up, overseasoning, and release of primary inoculum: Primary inoculum of Mycosphaerella graminicola. *Plant Pathology*, 60(2), 166–177. doi:10.1111/j.1365-3059.2010.02369.x.

Tabib Ghaffary, S. M., Robert, O., Laurent, V., Lonnet, P., Margalé, E., van der Lee, T. A. J., et al., 2011. Genetic analysis of resistance to Septoria tritici blotch in the French winter wheat cultivars

Balance and Apache. *Theoretical and Applied Genetics*, 123(5), 741–754. doi:10.1007/s00122-011-1623-7.

Talukder, R.K., 2000. A review of early warning systems practiced in selected Asian countries. Technical report [EWFIS Report-1]. Bangladesh. Ministry of Food. Food Planning and Monitoring Unit (FPMU)

Te Beest, D.E., Shaw, M.W., Pietravalle, S., and van den Bosch, F., 2009. A predictive model for early-warning of Septoria leaf blotch on winter wheat. *European Journal of Plant Pathology*, 124(3), 413–425. doi:10.1007/s10658-009-9428-0.

Thompson, S.K., 2012. « Stratified Sampling » In Sampling, Third Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

Van Maanen, A. and Xu, X.-M., 2003. Modelling plant disease epidemics. *European Journal of Plant Pathology*, 109(7), 669–682.

Verreet, J. A., Klink, H., and Hoffmann, G. M. (2000). Regional Monitoring for Disease Prediction and Optimization of Plant Protection Measures: The IPM Wheat Model. *Plant Disease*, 84(8), 816–826.

Wallach, D., Makowski, D., Jones, J.W., and Brun, F., 2014. Working with dynamic crop models. Methods, tools, and examples for agriculture and environment (Second ed.). Elsevier.

Whish, J.P.M., Herrmann, N.I., White, N.A., Moore, A.D., and Kriticos, D.J., 2015. Integrating pest population models with biophysical crop models to better represent the farming system. *Environmental Modelling & Software*, 72, 418–425. doi:10.1016/j.envsoft.2014.10.010

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. *Weed Research*, 14(6), 415–421.

Zhao, Y., Staudenmayer, J., Coull, B.A., and Wand, M.P., 2006. General Design Bayesian Generalized Linear Mixed Models. *Statistical Science*, 21(1), 35–51. doi:10.1214/088342306000000015

Annexes

Annexe 1: Correspondence between final leaf numbers and observed leaf numbers as a function of wheat stage on the observation date (Simonneau et al. 2009)

Observed leaf number	Zadoks stage (Zadoks et al., 1974) on the observation date	Corresponding final leaf number
1	Z31	4
1	Z32	3
1	Z37	2
1	Z39	1
2	Z31	5
2	Z32	4
2	Z37	3
2	Z39	2
3	Z31	6
3	Z32	5
3	Z37	4
3	Z39	3

Annexe 2: R code used to fit a frequentist model including site-year random effects and a covariable describing the level of SLB risk (model 4F)

This model was fitted with the `glmer()` function (available with the `lme4` package version 1.1–6). The SLB incidence data were included in a `data.frame` object called **TAB_glob**. The vector **feuille_m** includes the numbers of infected leaves in samples of 20 leaves. The vector **feuille_s** is the number of healthy leaves (**feuille_m** + **feuille_s** = 20). The variable **time_b** represents time in days (noted *t* in BMaterials and methods[^]), **risque** is a categorical variable indicating levels of SLB risk (high, medium, low) and **id_plot** is the site-year index variable.

In the `glmer()` function, the terms on the left of the sign «~» represent the data (**cbind(feuille_m,feuille_s)**) and the terms on the right describe the fixed effect (**time_b + risque**) and the random effect (**(1 + time_b|id_plot)**).

The number of infected leaves was assumed to follow a binomial probability distribution specified as **family = binomial**.

Model < – glmer(cbind(feuille_m, feuille_s) ~ time_b + risque + (1 + time_b | id_plot), family = binomial, data = TAB_glob)

Annexe 3: WinBUGS script used to fit a Bayesian model including site-year random effects and a covariable describing the level of SLB risk (model 4B)

Ndata is the total number of SLB incidence data, **Nsiteyear** is the number of site-years, **y[i]** is the i^{th} incidence, **theta[]** is a vector of random site-year effects, **rm[]** and **rf[]** are dummy variables indicating medium and high risk levels respectively, **p[]** is the probability that one leaf is infected, **alpha0**, **alpha1**, **betarm** and **betarf** are fixed effects, and **lambda** is a variable used to ensure that disease incidence is an increasing function of time (**time[]**).

```
model {  
  
  for (i in 1:Ndata){  
  
    y[i]~dbin(p[i],20)  
  
    lambda[i] <- max(0,(alpha1+theta[siteyear[i],2]))  
  
    logit(p[i]) <- alpha0+theta[siteyear[i],1]+lambda[i]*(time[i])+betarm*rm[i]+betarf*rf[i]  
  
  }  
  
  for (j in 1:Nsiteyear){  
  
    theta[j,1:2]~dmnorm(mu[1:2],tau.theta[1:2,1:2])  
  
  }  
  
  #prior  
  
  alpha0~dnorm(0,1.E-6)  
  alpha1~dnorm(0,1.E-6)  
  
  tau.theta[1:2,1:2] ~ dwish(R[1:2,1:2],2)  
  
  betarm~dnorm(0,1.E-6)  
  betarf~dnorm(0,1.E-6)  
  
}
```

Annexe 4: Estimated parameter values and 95 % confidence intervals (in brackets) obtained for the frequentist models 1F-3F, for the three leaves considered (leaves 1 to 3).

Parameters	Leaf 1		
	Model 1F	Model 2F	Model 3F
α_0	-7.223 (-7.403 ; -7.063)	-15.854 (-17.188 ; -14.52)	-8.083 (-8.282 ; -7.884)
α_1	0.110 (0.107 ; 0.113)	0.239 (0.214 ; 0.264)	0.112 (0.109 ; 0.116)
g_H			1.217 (1.108 ; 1.327)
g_M			0.616 (0.511 ; 0.721)

	Leaf 2		
	Model 1F	Model 2F	Model 3F
α_0	-6.083 (-6.201 ; -5.965)	-11.052 (-11.73 ; -10.373)	-6.551 (-6.685 ; -6.416)
α_1	0.094 (0.092 ; 0.096)	0.169 (0.155 ; 0.182)	0.096 (0.094 ; 0.099)
g_H			0.824 (0.747 ; 0.902)
g_M			0.203 (0.131 ; 0.276)

	Leaf 3		
	Model 1F	Model 2F	Model 3F
α_0	-5.121 (-5.212 ; -5.029)	-8.925 (-9.466 ; -8.383)	-5.495 (-5.598 ; -5.391)
α_1	0.075 (0.074 ; 0.077)	0.133 (0.123 ; 0.144)	0.076 (0.075 ; 0.078)
g_H			0.675 (0.613 ; 0.738)
g_M			0.250 (0.193 ; 0.308)

Annexe 5: Estimated posterior means of the parameters and 95 % credibility intervals (in brackets) obtained for Bayesian models 1B-3B, for the three leaves considered (leaves 1 to 3).

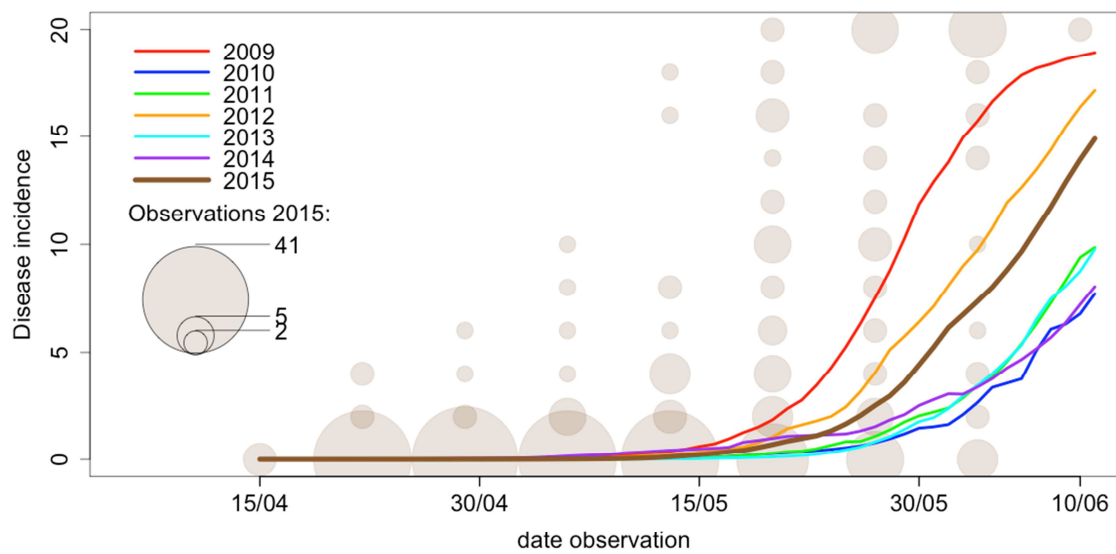
Parameters	Leaf 1		
	Model 1B	Model 2B	Model 3B
α_0	-7.235 (-7.408 ; -7.067)	-15.931 (-16.960 ; -14.86)	-8.085 (-8.286 ; -7.894)
α_1	0.110 (0.107 ; 0.113)	0.243 (0.223 ; 0.263)	0.112 (0.109 ; 0.115)
g_H			1.219 (1.111 ; 1.329)
g_M			0.618 (0.515 ; 0.722)

	Leaf 2		
	Model 1B	Model 2B	Model 3B
α_0	-6.084 (-6.203 ; -5.968)	-11.106 (-11.89 ; -10.43)	-6.552 (-6.688 ; -6.418)
α_1	0.094 (0.092 ; 0.096)	0.17 (0.158 ; 0.184)	0.096 (0.094 ; 0.098)
g_H			0.825 (0.749 ; 0.903)
g_M			0.204 (0.132 ; 0.276)

	Leaf 3		
	Model 1B	Model 2B	Model 3B
α_0	-5.121 (-5.212 ; -5.030)	-8.788 (-9.479 ; -8.196)	-5.495 (-5.599 ; -5.393)
α_1	0.075 (0.074 ; 0.077)	0.131 (0.119 ; 0.143)	0.076 (0.075 ; 0.078)
g_H			0.676 (0.614 ; 0.738)
g_M			0.251 (0.194 ; 0.301)

Annexe 6: Median regional disease incidence (%) estimated for leaf 2 from 2009 to 2015. The medians were calculated across the site-specific disease incidence dynamics estimated with model 4F fitted to 4617 SLB data collected from 2009 to 2015 in the Centre region.

Estimates obtained in 2015 (brown line) were derived from data collected before June 10th and extracted from the Vigiculture database. Light brown circles indicate observed data collected in the Centre region in 2015. Circle size is proportional to the number of data.



Annexe 7: Codes used to fit models

Frequentist models were fitted with the `glmer()` function of the `lme4` package version 1.1-6 of R (<https://www.r-project.org>). Bayesian models were fitted with winBUGS software (<http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>).

The 10septo frequentist model

The STB incidence data were included in a `data.frame` object called **TAB_septo**. The vector **leaves_m** is the number of infected leaves in a sample of 20 leaves. The vector **leaves_s** is the number of healthy leaves (**leaves_m** + **leaves_s** = 20). The variable **time** is the time in days (called *t* in the “General framework” paragraph of the “Materials & methods”), **risk** and **region** are categorical variables indicating the level of STB risk (high, medium, low) and the region (Champagne-Ardenne and Centre), respectively and **id_unique** is the site-year index variable.

In the `glmer()` function, the terms on the left of the sign “~” represent the data (**cbind(leaves.m,leaves.s)**) and the terms on the right describe the fixed effect (**time * region + risk**) and the random effect (**(1+time|id_unique)**).

The number of infected leaves was assumed to follow a binomial probability distribution specified as **family=binomial**.

```
10septo <- glmer(cbind(leaves.m,leaves.s)~time*region+risk + (1+time|id_unique),
family=binomial, data=TAB_septo)
```

The 10septo Bayesian model

Ndata is the total number of STB incidence data, **Nsiteyear** is the number of site-years, **y[i]** is the *i*th incidence, **theta[]** is a vector of random site-year effects, **rm[]** and **rf[]** are dummy variables indicating medium and high risk levels, respectively, **regChamp []** is a dummy variable indicating the Champagne-Ardenne region, **p[]** is the probability that one leaf is infected, **alpha0**, **alpha1**, **betarm**, **betarf**, **betareg0** and **betareg1** are fixed effects, and **lambda** is a variable used to ensure that disease incidence is an increasing function of time (**time[]**).

```
model {
```



```
for (i in 1:Ndata){
y[i]~dbin(p[i],20)
lambda[i] <- alpha1+theta[siteyear[i],2]+betareg1*regChamp[i]
logit(p[i]) <-
alpha0+theta[siteyear[i],1]+lambda[i]*(time[i])+betarm*rm[i]+betarf*rf[i]+betareg0*regChamp[i]
}

for (j in 1:Nsiteyear) {
theta[j,1:2]~dmnorm(mu[1:2],tau.theta[1:2,1:2])
}
#prior
alpha0~dnorm(0,1.E-4)
alpha1~dnorm(0,1.E-4)
tau.theta[1:2,1:2] ~ dwish(R[1:2,1:2],2)
betarm~dnorm(0,1.E-4)
betarf~dnorm(0,1.E-4)
betareg0~dnorm(0,1.E-4)
betareg1~dnorm(0,1.E-4)
}
```

The 6vine frequentist model

The proportions of vines infected were included in a data.frame object called **TAB_vine**. The vector **cep_infected** is the number of vines infected. The vector **cep_safe** is the number of healthy vines. The variable **time** is the time in days (called *t* in “General framework” paragraph of the “Materials & methods”), **id_site_year** is the site-year index variable and **year** is the year index variable.

In the `glmer()` function, the terms on the left of the sign “~” represent the data (**cbind(cep_infected,cep_safe)**) and the terms on the right describe the fixed effect (**(time)**) and the random effects (**((1+time|id_site_year)+(1+time|year))**).

The number of infected vines was assumed to follow a binomial probability distribution specified as **family=binomial**.

```
6vine<- glmer(cbind(cep_infected,cep_safe)~time + (1+time|id_site_year)+ (1+time|year),  
family=binomial, data=TAB_vine)
```

The 6vine Bayesian model

Ndata is the proportion of vines infected, **Nsiteyear** is the number of site-years, **Nyear** is the number of years, **y[i]** is the i^{th} proportion, **n[i]** is the i^{th} total number of vines observed, **theta0** and **beta0** are vectors of random year and site-year effects, respectively, **p[]** is the probability that one vine is infected, **alpha0** and **alpha1** are fixed effects; the proportion of vines infected is a function of time (**time[]**).

```
model {  
for (i in 1:Ndata)  
{  
y[i]~dbin(p[i],n[i])  
logit(p[i]) <- alpha0+theta0[year[i]]+beta0[siteyear[i]]+alpha1*time[i]  
}  
for (j in 1:Nyear) {  
theta0[j]~dnorm(0.0,tau.theta0)  
}  
for (k in 1:Nsiteyear) {  
beta0[k]~dnorm(0.0,tau.beta0)  
}  
#prior  
alpha0~dnorm(0,1.0E-6)  
alpha1~dnorm(0,1.0E-6)  
tau.beta0~dgamma(0.001,0.001)  
tau.theta0~dgamma(0.001,0.001)  
}
```

The 5Sigatoka frequentist model

The model was fitted with the `glmer()` function (available with the `lme4` package version 1.1-6). The probabilities of SED exceeding a critical threshold equal to 250 were included in a `data.frame` object called **TAB_banana**. The vector **threshold_250** corresponds to the probability of SED exceeding a critical threshold equal to 250. The variable **month** is the time in months (called *t* in the “General framework” paragraph of the “Materials & methods”), **site_year** is the site-year index variable and **annee** is the year index variable.

In the `glmer()` function, the terms on the left of the sign “~” represent the data (**`cbind(threshold_250,(1- threshold _250))`**) and the terms on the right describe the fixed effect (**`month`**) and the random effects (**`1|site_year`**).

The probability of SED exceeding a critical threshold equal to 250 was assumed to follow a binomial probability distribution specified as **`family=binomial`**.

```
5Sigatoka <- glmer(cbind(threshold_250,(1- TAB_threshold _250))~month+ (1|site_year),  
family=binomial, data=TAB_banana)
```

The 5Sigatoka Bayesian model

Ndata is the probability of SED exceeding a critical threshold equal to 250, **Nsiteyear** is the number of site-years, **y[i]** is the *i*th probability, **theta[]** is a vector of random site-year effects , **p[]** is the probability of SED exceeding a critical threshold equal to 250, **alpha8**, and **delta[]** are fixed effects, the probability of SED exceeding a critical threshold equal to 250 is a function of month (**month[]**).

```
model {  
  delta[8]<-0  
  for (i in 1:Ndata){  
    y[i]~dbern(p[i])  
    logit(p[i]) <- alpha8+delta[month[i]]*(1-equals(month[i],8)) + theta[siteyear[i]]  }  
  
  for (j in 1:Nsiteyear) {  
    theta[j]~dnorm(0.0,tau.theta)  
  }  
}
```

```
#prior
for(k in 1:7) {
  delta[k]~dnorm(0,1.E-1)
}
for(h in 9:12) {
  delta[h]~dnorm(0,1.E-1)
}
tau.theta ~ dgamma(0.001,0.001)
alpha8~dnorm(0,1.0E-1)
}
```

The 5weevils frequentist model

The model was fitted with the `glmer()` function (available with the `lme4` package version 1.1-6). The number of trapped weevils was included in a data.frame object called **TAB_weevils**. The vector **capture** is the number of trapped weevils. The variable **month** represents time in months (called *t* in the “General framework” paragraph of the “Materials & methods”), **site_year** is the site-year index variable.

In the `glmer()` function, the terms on the left of the sign “~” represent the data (**capture**) and the terms on the right describe the fixed effect (**month**) and the random effects (**1|site_year**).

The number of trapped weevils was assumed to follow a Poisson probability distribution specified as **family=poisson**.

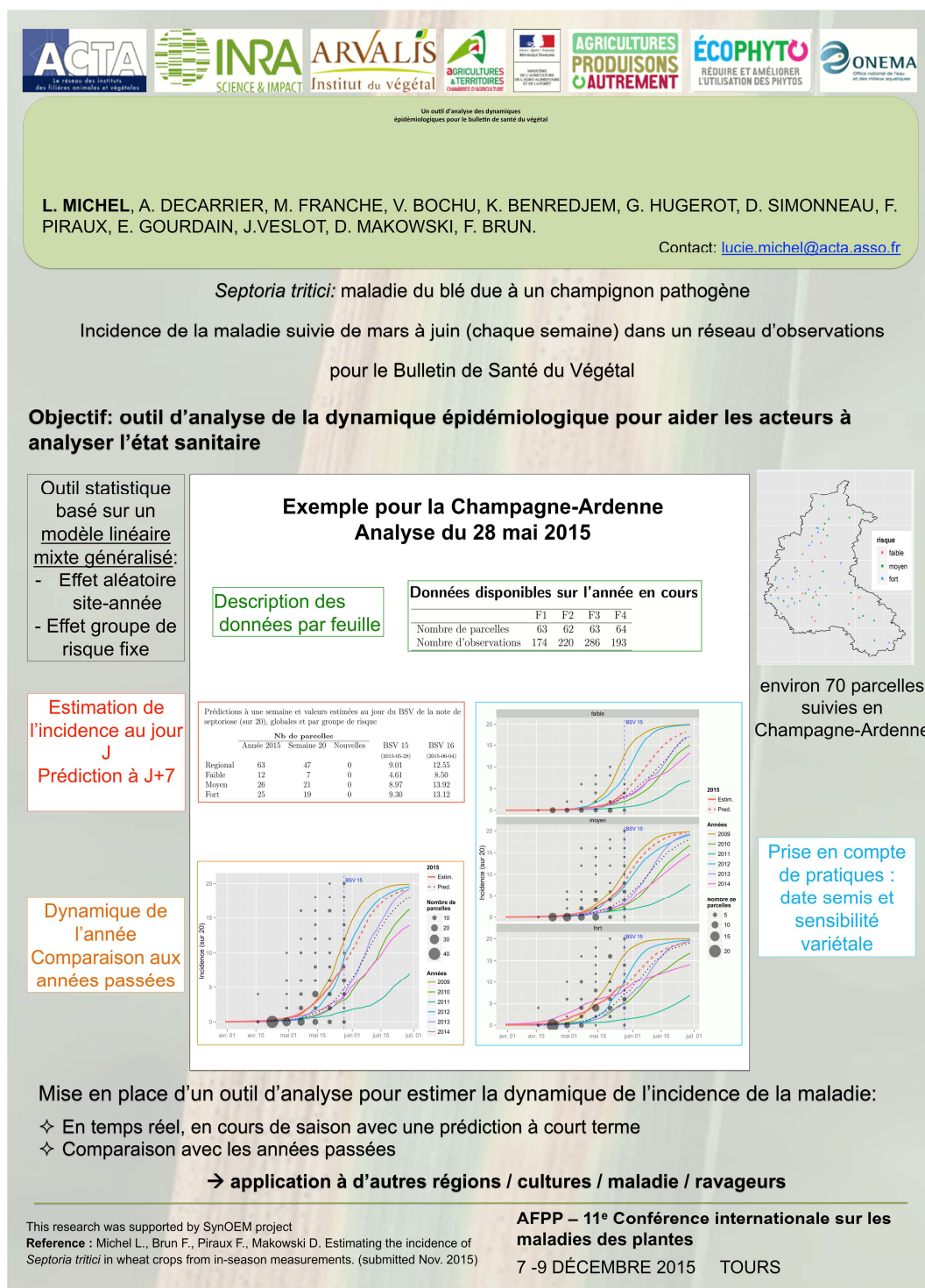
```
5weevils <- glmer(capture ~ month + (1|site_year) , family = poisson(),data=TAB_weevils)
```

The 5weevils Bayesian model

Ndata is the number of trapped weevils, **Nsiteyear** is the number of site-years, **y[i]** is the *i*th number of trapped weevils, **theta[]** is a vector of random site-year effects, **p[]** is the number of trapped weevils, **alpha1**, and **delta[]** are fixed effects, the number of trapped weevils is a function of month (**month[]**).

```
model {  
delta[1]<-0  
for (i in 1:Ndata){  
y[i]~dpois(mu[i])  
log(mu[i]) <- alpha1+delta[mois[i]]*(1-equals(mois[i],1)) + theta[siteyear[i]] }  
  
for (j in 1:Nsiteyear) {  
theta[j]~dnorm(0.0,tau.theta)  
}  
#prior  
for(k in 2:12) {  
delta[k]~dnorm(0,1.E-1)  
}  
tau.theta ~ dgamma(0.001,0.001)  
alpha1~dnorm(0,1.0E-1)  
}
```

Annexe 8 : Poster présenté à l'AFPP- onzième conférence internationale sur les maladies des plantes Tours – 7 au 9 décembre 2015



Abstract

Pest and disease monitoring has been strengthened in France since 2009, following the introduction of the National Ecophyto plan. It is based on monitoring a network of cultivated plots. Several diseases and pests are monitored on these plots following harmonized protocols. Observations are used to develop plant health bulletins, but these data are underused, their current use is largely restricted to descriptive statistics due to lack of time and appropriate tools.

The main objective of this PhD research is to develop an operational statistical approach to analyse epidemiological dynamics based on this pests and diseases system of monitoring.

An approach has been developed on *Septoria* leaf blotch in wheat in Champagne-Ardenne. It is based on a generalised linear mixed model taking into account random site-year effects and a risk rating fixed effects. Three risk ratings (high, medium, low) are defined by a simple classification rule defined by the Champagne-Ardenne extension service based on sowing date and varietal resistance. This model can be easily adjusted to field monitoring of pests and diseases in this region, and it can be updated weekly, as new data become available. The model estimates the dynamics of *Septoria* incidence at a regional scale taking into account field risk ratings. These dynamics can be presented in graphical form and can be easily integrated in weekly reports for the plant health bulletins. The statistical model was evaluated with field observations between 2009 and 2014. Its practical use was tested in collaboration with the Champagne-Ardenne extension services in 2015.

The approach developed for *Septoria* leaf blotch was generalised and applied to other regions and other diseases and pests. The statistical model was modified for use in several other French regions. Variants have also been proposed to analyse observations of downy mildew on grapevine in the Midi-Pyrénées region, or observations of Yellow Sigatoka on banana and counts of weevils on sweet potato in Guadeloupe. These variants differ by the type of random effects and fixed effects included in the model. The best performing model was selected using several criteria: Akaike information criterion, Root mean square error, and residuals analysis. In each case study, each model was compared to its Bayesian version, developed using uninformative prior distributions. Results show that classical and Bayesian approaches generate very similar results, but it also shows the advantage of Bayesian statistics in analyzing the uncertainty associated with site-specific predictions.

The third part of this PhD research shows how these models could also be used to optimise the size of the network of monitored fields. Currently, fields monitored for pests and diseases in Champagne-Ardenne are observed at least weekly and these observations involve significant costs. Predictions of our statistical models allow for stratification field sampling and make it possible to reduce its size over the growing season.

Finally, the last part of this thesis addresses the implementation of this statistical approach to agricultural extension services in charge of developing the plant health bulletins. This analysis was based on discussions with national and regional partners such as agricultural technical institutes, the Chamber of Agriculture, and the French agricultural services department, which took place early on in the project, and was completed by a survey of the people involved in producing the plant health bulletins.

Keywords

disease monitoring program, observation network, statistical modeling, predictions, sampling, stratification

Résumé

La Surveillance Biologique du Territoire (SBT) s'est renforcée en France depuis 2009, à la suite du Plan Écophyto. Elle repose sur le suivi d'un réseau de parcelles cultivées. Plusieurs maladies et ravageurs sont observés sur ces parcelles selon des protocoles harmonisés. Les observations sont utilisées pour élaborer le Bulletin de Santé du Végétal (BSV), mais la valorisation actuelle de ces données reste limitée et repose essentiellement sur des analyses descriptives faute de temps et d'outils adaptés.

Le principal objectif de ces travaux de thèse est de développer une démarche statistique opérationnelle pour analyser des dynamiques épidémiologiques à partir des observations de terrain collectées dans le cadre de la SBT.

La démarche a été mise au point sur un cas d'étude concernant la septoriose du blé en Champagne-Ardenne. Elle est basée sur un modèle linéaire mixte généralisé prenant en compte un effet site-année aléatoire et un effet niveau de risque fixe. Trois niveaux de risque (fort, moyen ou faible) sont définis en appliquant une règle de décision définie par les experts de la région tenant compte de la date de semis et de la résistance variétale. Ce modèle peut être facilement ajusté aux observations collectées dans le réseau de parcelles suivies dans cette région, et peut être mis à jour semaine après semaine, à chaque fois que de nouvelles données sont disponibles. Le modèle permet d'estimer la dynamique de l'incidence de la septoriose du blé à l'échelle régionale en tenant compte du niveau de risque des parcelles. Ces dynamiques peuvent être présentées sous forme de graphiques et être facilement intégrées dans les rapports hebdomadaires utilisés lors de l'élaboration du BSV. Le modèle statistique a été évalué sur des observations collectées entre 2009 et 2014. Son utilisation pratique a été testée en collaboration avec la chambre d'agriculture de Champagne-Ardenne en 2015.

La démarche mise au point sur la septoriose a été généralisée et appliquée à d'autres régions et d'autres maladies ou ravageurs. Le modèle statistique a été modifié pour pouvoir être appliqué à plusieurs régions françaises. Des variantes ont également été proposées pour pouvoir analyser des observations du mildiou de la vigne en région Midi-Pyrénées, des observations de cercosporiose jaune du bananier et des comptages de charançon sur patate douce en Guadeloupe. Ces variantes diffèrent selon le type d'effets aléatoires et d'effets fixes introduits dans le modèle. La sélection du modèle le plus adapté à la situation considérée a été réalisée en utilisant plusieurs critères : Akaike information criterion, Root mean square error, analyse des résidus. Dans chacune des situations considérées, chaque modèle a été comparé à sa version bayésienne, développée en utilisant des distributions a priori peu informatives. Les résultats montrent que les versions classiques et bayésiennes donnent des résultats très proches, mais montrent également l'intérêt du Bayésien pour analyser l'incertitude des prédictions.

Dans la troisième partie, nous avons montré comment ces modèles pouvaient également être utilisés pour optimiser la taille de l'échantillon du réseau de parcelles de SBT. Actuellement, les parcelles du réseau d'épidémiosurveillance en Champagne-Ardenne sont observées environ une fois par semaine et ces observations engendrent des coûts importants. Les prédictions de nos modèles statistiques permettent de stratifier l'échantillon de parcelles et de réduire sa taille en cours de saison.

Finalement, dans la dernière partie de cette thèse, je discute la mise en œuvre opérationnelle de cette démarche statistique auprès des acteurs chargés de l'élaboration du BSV. Cette discussion repose sur les échanges avec les partenaires au niveau national et régional (Instituts Techniques Agricoles, Chambre d'agriculture, services de l'état) qui ont eu lieu dès le début du projet, mais aussi sur une enquête réalisée auprès des acteurs du BSV.

Mots-clés

Surveillance Biologique du Territoire, réseau d'observation, modélisation statistique, prédictions, échantillonnage, stratification